Quantenklassische Hybridbeschreibung von Solvatisierungseffekten: Kombination von Dichtefunktionaltheorie mit polarisierbaren molekülmechanischen Kraftfeldern

Magnus Schwörer

München 2016

Quantenklassische Hybridbeschreibung von Solvatisierungseffekten: Kombination von Dichtefunktionaltheorie mit polarisierbaren molekülmechanischen Kraftfeldern

Magnus Schwörer

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Magnus Schwörer
aus München

München, den 23.03.2016

Gutachter: Prof. Dr. Paul Tavan
Zweitgutachterin: Prof. Dr. Regina de Vivie-Riedle
Tag der mündlichen Prüfung: 03.05.2016

Kurzfassung

Eine aussagekräftige theoretische Beschreibung des Infrarot (IR) -Schwingungsspektrums eines Biomoleküls in seiner nativen Umgebung durch Molekulardynamik (MD) -Simulationen benötigt hinreichend genaue Modelle sowohl für das Biomolekül, als auch für das umgebende Lösungsmittel. Die quantenmechanische Dichtefunktionaltheorie (DFT) stellt solche genauen Modelle zur Verfügung, zieht aber hohen Rechenaufwand nach sich. Daher ist dieser Ansatz nicht zur Simulation der MD ausgedehnter Biomolekül-Lösungsmittel-Komplexe einsetzbar. Solche Systeme können effizient mit polarisierbaren molekülmechanischen (PMM) Kraftfeldern behandelt werden, die jedoch nicht die zur Berechnung von IR-Spektren nötige Genauigkeit liefern.

Einen Ausweg aus dem skizzierten Dilemma bieten Hybridverfahren, die einen relevanten Teil eines Simulationssystems mit DFT, und die ausgedehnte Lösungsmittelumgebung mit einem (P)MM-Kraftfeld beschreiben. Im Rahmen dieser Arbeit wird, ausgehend von einer DFT/MM-Hybridmethode [Eichinger *et al.*, J. Chem. Phys. **110**, 10452-10467 (1999)], ein genaues und hocheffizientes DFT/PMM-Rechenverfahren entwickelt /2,4,6,7/, das der wissenschaftlichen Öffentlichkeit nun in Form des auf Großrechnern einsetzbaren Programmpakets IPHIGENIE/CPMD zur Verfügung steht.

Die neue DFT/PMM-Methode fußt auf der optimalen Integration des DFT-Fragments in die „schnelle strukturadaptierte Multipolmethode" (SAMM) zur effizienten approximativen Berechnung der Wechselwirkungen zwischen den mit gitterbasierter DFT bzw. mit PMM beschriebenen Subsystemen. Dies erlaubt stabile Hamilton'sche MD-Simulationen /2/ sowie die Steigerung der Performanz (d.h. dem Produkt aus Genauigkeit und Recheneffizienz) um mehr als eine Größenordnung /4/. Die eingeführte explizite Modellierung der elektronischen Polarisierbarkeit im PMM-Subsystem durch induzierbare Gauß'sche Dipole ermöglicht die Verwendung wesentlich genauerer PMM-Lösungsmittelmodelle /2/. Ein effizientes Abtastens von Peptidkonformationen mit DFT/PMM-MD kann mit einer generalisierten Ensemblemethode erfolgen /7/.

Durch die Entwicklung eines Gauß'schen polarisierbaren Sechspunktmodells (GP6P) für Wasser und die Parametrisierung der Modellpotentiale für van der Waals-Wechselwirkungen zwischen GP6P-Molekülen und der Amidgruppe (AG) von N-Methyl-Acetamid (NMA) wird ein DFT/PMM-Modell für (Poly-)Peptide und Proteine in wässriger Lösung konstruiert /6/. Das neue GP6P-Modell kann die Eigenschaften von flüssigem Wasser mit guter Qualität beschreiben. Ferner können die mit DFT/PMM-MD berechneten IR-Spektren eines in GP6P gelösten DFT-Modells von NMA die experimentelle Evidenz mit hervorragender Genauigkeit reproduzieren /6/. Somit ist nun ein hocheffizientes und ausgereiftes DFT/PMM-MD-Verfahren zur genauen Berechnung der Konformationslandschaften und IR-Schwingungsspektren von in Wasser gelösten Proteinen verfügbar.

Abstract

A meaningful theoretical description of the infrared (IR) spectrum of a biomolecule in its native environment by molecular dynamics (MD) simulations requires adequately accurate models both for the biomolecule and for its solvent environment. The quantum mechanical density functional theory (DFT) provides such accurate models, but entails high computational effort. Therefore, this approach is not suited for the simulation of the MD of extended biomolecule-solvent-complexes. Such systems can be handled efficiently by polarizable molecular mechanics (PMM) force fields, which, however, do not provide the accuracy required for the computation of IR spectra.

The sketched dilemma is resolved by hybrid approaches, which describe a relevant part of a simulation system by DFT, and the extended solvent environment by a (P)MM force field. Based on a DFT/MM hybrid method [Eichinger *et al.*, J. Chem. Phys. **110**, 10452-10467 (1999)], an accurate and highly efficient DFT/PMM approach is developed in this thesis /2,4,6,7/. Its implementation in the program package IPHIGENIE/CPMD is suitable for high-performance computing applications and available to the scientific community.

The new DFT/PMM method is based on the optimal integration of the DFT fragment into the "structure-adapted fast multipole method" (SAMM) for the efficient approximative computation of interactions between the subsystems described by grid-based DFT and PMM, respectively. It enables stable, Hamiltonian MD simulations /2/, and increases the performance (i.e. accuracy times efficiency) by more than one order of magnitude /4/. The explicit modeling of electronic polarizability in the PMM subsystem by induced Gaussian dipoles allows the use of much more accurate PMM solvent models /2/. The efficiency of peptide conformational sampling with DFT/PMM-MD is increased by applying a generalized ensemble method /7/.

By constructing a Gaussian polarizable six-point (GP6P) model for water and by parametrizing the model potentials for van der Waals interactions between GP6P molecules and the amide group (AG) of N-Methyl-Acetamide (NMA), a DFT/PMM model for (poly-)peptides and proteins in aqueous solution is developed /6/. The new GP6P model can describe the properties of liquid water with good quality. Furthermore, the IR spectra of a DFT model of NMA solvated in GP6P, which were calculated by DFT/PMM-MD, can reproduce the experimental evidence with excellent quality /6/. Thus, a highly efficient and mature DFT/PMM-MD approach for the accurate computation of conformational landscapes and IR spectra of proteins in aqueous solution is now available.

Verzeichnis der im Rahmen meiner Arbeit in der Arbeitsgruppe für theoretische molekulare Biophysik am Lehrstuhl für BioMolekulare Optik der Ludwig-Maximilians-Universität München entstandenen Publikationen

/1/ K. Lorenzen, M. Schwörer, P. Tröster, S. Mates und P. Tavan. Optimizing the accuracy and efficiency of fast hierarchical multipole expansions for MD simulations. *J. Chem. Theory Comput.* **8**, 3628-3636 (2012).

/2/ **M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan und G. Mathias. Coupling DFT to polarizable force fields for efficient and accurate Hamiltonian molecular dynamics simulations. *J. Chem. Phys.* 138, 244103 (2013).**

/3/ P. Tröster, K. Lorenzen, M. Schwörer und P. Tavan. Polarizable water models from mixed computational and empirical optimization. *J. Phys. Chem. B* **117**, 9486-9500 (2013).

/4/ **M. Schwörer, K. Lorenzen, G. Mathias und P. Tavan. Utilizing Fast Multipole Expansions for Efficient and Accurate Quantum-Classical Molecular Dynamics Simulations. *J. Chem. Phys.* 142, 104108 (2015).**

/5/ M. Schwörer und G. Mathias. Modeling Biomolecular Solvation Effects by Hybrid QM/MM Methods. In G. Sutmann, J. Grotendorst, G. Gompper, D. Marx (Herausgeber): *Computational Trends in Solvation and Transport in Liquids*. Schriften des Forschungszentrums Jülich, IAS Series 28, 315-332 (2015).

/6/ **M. Schwörer, C. Wichmann und P. Tavan. A Polarizable QM/MM Approach to the Molecular Dynamics of Amide Groups Solvated in Water. *J. Chem. Phys.* 144, 114504 (2016).**

/7/ **M. Schwörer, C. Wichmann, E. Gawehn und G. Mathias. Simulated Solute Tempering in Fully Polarizable Hybrid QM/MM Molecular Dynamics Simulations. *J. Chem. Theory Comput.* 12, 992-999 (2016).**

Die durch Fettdruck hervorgehobenen Arbeiten sind in den Text der Dissertation eingearbeitet und dort nachgedruckt.

Inhaltsverzeichnis

1 Einleitung

Proteine stellen, neben der die Erbinformation kodierenden Desoxyribonukleinsäure (engl. *deoxyribonucleic acid*, DNA) und dem Wasser, in dem sich biochemische Prozesse abspielen, die wichtigste Klasse von Biomolekülen dar [1]. So katalysieren Enzyme biochemische Prozesse wie die Vervielfältigung der DNA, Kanal- und Pumpproteine steuern die Chemikalienkonzentration im wässrigen Zellinneren, Sinneswahrnehmungen und Photosyntheseprozesse werden durch Proteine organisiert. Auch die Proteinsynthese selbst wird durch Proteine bewerkstelligt [1].

Die Funktion eines Proteins beruht auf seiner spezifischen dreidimensionalen Form [1] und seiner Konformationsdynamik [2, 3]. Kleinste strukturelle Fehler oder Schäden können bereits die Proteinfunktion stören oder zerstören [4, 5], neurodegenerative Krankheiten wie Alzheimer oder Parkinson werden durch fehlgefaltete Proteine ausgelöst [6–8]. Eine genaue Kenntnis der Struktur und Funktionsdynamik von Proteinen ist daher eine wichtige Grundlage für das Verstehen von biologischen Prozessen und für die Diagnose und Behandlung einer Vielzahl von Krankheiten.

Die strukturellen und dynamischen Eigenschaften eines Proteins werden durch komplexe intramolekulare Wechselwirkungen zwischen seinen Grundbausteinen, den Aminosäuren, und durch seine Umgebung festgelegt [2]. Im Allgemeinen kann ein Protein seine native Form nur im physiologischen Milieu der Zelle, das durch eine komplex strukturierte wässrige Lösung von Biomolekülen und Ionen charakterisiert ist, annehmen und behalten. Hier liefern insbesondere die polaren Wassermoleküle einen großen entropischen Beitrag zur Stabilität von Proteinen [2], und tragen weiterhin durch spezifische elektrostatische Wechselwirkungen zur lokalen Strukturbildung bei [9]. Ferner beeinflussen die durch die Wassermoleküle erzeugten elektrischen Felder die elektronischen Eigenschaften und die molekularen Schwingungsmuster von Peptiden und Proteinen [10].

Eine Vielzahl experimenteller biophysikalischer Techniken steht für die Aufklärung der Struktur und Funktionsdynamik von Proteinen zur Verfügung [11]. Theoretische biophysikalische Methoden, wie die Molekulardynamik (MD) - Simulationen [12–15] helfen, die experimentellen Befunde zu interpretieren und zu ergänzen [16–22]. So können z.B. molekulare Strukturdaten aus Röntgenbeugungs-, Neutronenstrahlungs- oder Kernspinresonanzexperimenten durch MD-Techniken verifiziert und verfeinert werden [16, 17]. Das komplexe Schwingungsspektrum eines (Bio-)Moleküls im Infrarot (IR) - Bereich, das dynamische Informationen u.a. über dessen chemischen Zustand [23–25] und über die Natur und Struktur seiner Umgebung [10, 26] liefert, kann mit Hilfe von MD-Simulationen interpretiert werden [19, 21, 27].

Anhand vereinfachter biophysikalischer Modelle können die für eine Beobachtung verantwortlichen mikroskopischen physikalischen Effekte identifiziert werden, oder, bei hinrei-

chender Modellqualität, Vorhersagen für experimentell nicht oder nur schwer zugängliche Observablen gemacht werden [13]. Neben den für die Untersuchung von Proteinen relevanten Längen- und Zeitskalen, die sich über mehrere Größenordnungen erstrecken [3], stellt bei MD-Simulationen vor allem die hinreichend genaue, aber trotzdem effiziente Beschreibung der physikalischen Eigenschaften von Protein-Wassergemischen eine große Herausforderung dar [27–30].

Ein ausgezeichneter Kompromiss aus Genauigkeit und Recheneffizienz kann durch die erstmals von Arieh Warshel und Michael Levitt im Jahre 1976 vorgeschlagenen hybriden Simulationsmodelle erreicht werden [31], die für die Behandlung eines ausgedehnten Biomolekül-Lösungsmittel-Komplexes eine genaue, aber rechenaufwändige quantenmechanische (QM) Beschreibung eines — z.B. chemisch reaktiven oder spektroskopisch interessanten — molekularen Subsystems mit einer ungenaueren, aber sehr effizienten Behandlung des umgebenden Lösungsmittelsystems durch ein sog. molekülmechanisches (MM) Kraftfeld kombinieren. Dafür wurden die beiden genannten Autoren, zusammen mit Martin Karplus, mit dem Chemie-Nobelpreis des Jahres 2013 ausgezeichnet [32].

Aufbauend auf dieser Idee entwickelten Eichinger *et al.* eine theoretische Methode [33], die eine Beschreibung des QM-Subsystems mittels Dichtefunktionaltheorie (DFT) ermöglichte, so die zur Berechnung von IR-Spektren erforderliche Genauigkeit lieferte und speziell auf Biomoleküle in deren nativer Umgebung zugeschnitten war. Eine Vielzahl von Anwendungen (siehe z.B. Literaturzitate [21], [25], oder [34]) zeigte, dass dieses DFT/MM-Verfahren eine zufriedenstellend genaue Berechnung solcher Schwingungsspektren ermöglichte. Die Vernachlässigung von elektronischen Polarisationseffekten im MM-Fragment stellte sich allerdings als zu grobe Vereinfachung bei der Beschreibung der Lösungsmitteleffekte auf das DFT-Fragment heraus [35–37]. Des Weiteren waren ausgedehnte, stabile MD-Simulationen mit DFT/MM-Modellen aufgrund weiterer verwendeter Näherungen nur eingeschränkt möglich.

Das Ziel der vorliegenden Arbeit war daher die Weiterentwicklung dieser DFT/MM-Technik [33] hinsichtlich ihrer Eignung für stabile MD-Simulationen und der Genauigkeit der verwendbaren Lösungsmittelmodelle. Besondere Aufmerksamkeit sollte dabei einer physikalisch korrekten, genauen und hocheffizienten Behandlung der Wechselwirkungen zwischen den Subsystemen gewidmet werden. Die Einführung polarisierbarer Freiheitsgrade im MM-Fragment sollte es erlauben, Polarisationseffekte auch dort zu berücksichtigen.

Zur Beschreibung einer wässrigen Lösungsmittelumgebung wurde ein polarisierbares MM (PMM) Wassermodell speziell für DFT/PMM-Anwendungen entwickelt, das die Solvatstruktur um das DFT-Subsystem, und die durch das Wasser hervorgerufenen Solvatisierungseffekte korrekt beschreiben kann. Die neue DFT/PMM-Methode soll so eine wesentlich genauere und effizientere Berechnung der IR-Spektren von in Wasser gelösten Molekülen erlauben.

Bevor die eben angesprochenen theoretischen Methoden jedoch genauer erklärt werden, werden im Folgenden zunächst einige grundlegende biochemische und physikalische Eigenschaften von Wasser und von Proteinen skizziert.

1.1 Proteine in wässriger Lösung

Die Grundbausteine der den Proteinen zugrundeliegenden Polypeptidketten sind die zwanzig α-Aminosäuren, die bei der *Translation* im Ribosom in einer für jedes Protein wohldefinierten Reihenfolge, der in der DNA kodierten *Primärstruktur*, als zusammenhängender Rückgratstrang polymerisiert werden [1]. Anschließend bilden sich lokale geordnete Muster (*Sekundärstrukturen*), die dann zusammen mit ungeordneten, flexiblen Bereichen (engl. *random coils*) die dreidimensionale *Tertiärstruktur* eines Proteins, bzw. einer Untereinheit eines Proteinkomplexes, formen [1]. Die nicht-kovalente Verbindung mehrerer Untereinheiten zu einem Proteinkomplex bezeichnet man auch als *Quartärstruktur* [1]. Der Faltungsprozess hängt wesentlich von der Umgebung des Proteins ab [38, 39], die *in situ* hauptsächlich durch wässrige Lösung gegeben ist.

1.1.1 Grundlegende Eigenschaften von Wasser

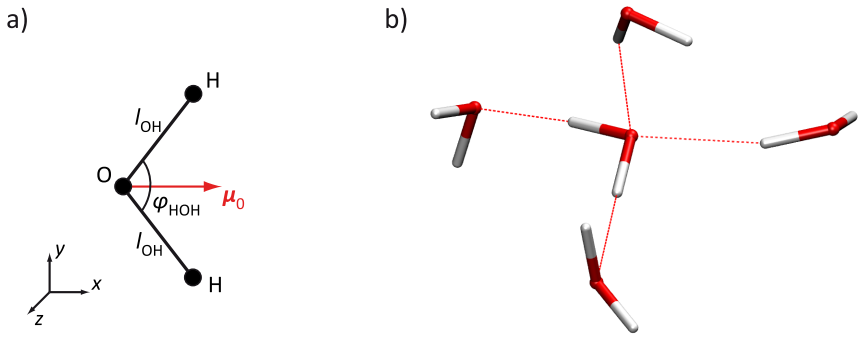


Abbildung 1.1: a) Geometrie eines H_2O-Moleküls in der Gasphase, dessen statisches Dipolmoment $\boldsymbol{\mu}_0$ zeigt in Pfeilrichtung. b) Beispiel für ein Wasserstoffbrückennetzwerk in flüssiger Phase: das zentrale Wassermolekül bildet vier Wasserstoffbrücken (gestrichelt) zu umgebenden Wassermolekülen aus.

Wasser, die Grundlage organischen Lebens [40], spielt eine besondere Rolle für biologische Prozesse [9]. Der Grund sind seine besonderen mikroskopischen und makroskopischen Eigenschaften [9, 41, 42]. Zum Beispiel weist flüssiges Wasser bei Raumtemperatur eine hohe Dielektrizitätskonstante von $\epsilon_{H_2O} \approx 78$ auf [43], d.h. die effektive Reichweite elektrostatischer Wechselwirkungen ist im Wasser gegenüber dem Vakuum auf $1/\epsilon_{H_2O}$ verringert. Weiterhin zeigt flüssiges Wasser, neben vieler anderer ungewöhnlicher Eigenschaften, eine Anomalie der Dichte, die bei 4° C ein Maximum aufweist [44].

Wie in Abbildung 1.1a gezeigt ist, sind die zwei Wasserstoffatome (H) und das Sauerstoffatom (O) in der Gasphase mit einer O–H-Bindungslänge $l_{OH}^{gas} = 0{,}9572\,\text{Å}$ und unter einem H–O–H Bindungswinkel $\varphi_{HOH}^{gas} = 104{,}52°$ angeordnet [45, 46]. Durch die unterschiedlichen atomaren Elektronegativitäten ergibt sich ein relativ starkes statisches Dipolmoment von $|\boldsymbol{\mu}_0| = 1{,}855\,\text{D}$ [46] in x-Richtung (roter Pfeil). Ferner hat das Wassermolekül ein großes Quadrupolmoment [47]. Diese elektrostatische Signatur generiert spezifisch strukturierte elektrische Felder.

Ein auf ein Molekül wirkendes äußeres elektrisches Feld hat eine Verschiebung der Elektronenverteilung relativ zu den Atomkernen zur Folge. Diese *elektronische Polarisierung* bewirkt u.a. eine Änderung von Geometrie und Dipolmoment. In wässriger Lösung wächst die mittlere Bindungslänge auf $l_{\mathrm{OH}}^{\mathrm{sol}} = 0{,}968\,\mathrm{\AA}$ und der Bindungswinkel auf $\varphi_{\mathrm{HOH}}^{\mathrm{sol}} = 105{,}3°$ [48, 49]. DFT-Rechnung ergeben, dass der Nettoeffekt dieser Geometrieänderung auf den statischen Dipol klein ist (siehe Abb. 2A in Ref. [50]).

Der durch ein äußeres elektrisches Feld in einem isolierten Wassermolekül induzierte Dipol $\boldsymbol{\mu}^{\mathrm{i}}$ ist in erster Näherung linear von diesem Feld abhängig; die entsprechenden linearen Koeffizienten $\alpha_{xx} = 1{,}47\,\mathrm{\AA}^3$, $\alpha_{yy} = 1{,}53\,\mathrm{\AA}^3$ und $\alpha_{zz} = 1{,}42\,\mathrm{\AA}^3$ sind annähernd isotrop [51] und ändern sich beim Übergang in die flüssige Phase nur unwesentlich [50]. In flüssigem Wasser wirken die elektrischen Felder im Mittel parallel zur x-Achse des in Abbildung 1.1a definierten molekülfesten Koordinatensystems, sodass der Gesamtdipol des einzelnen Wassermoleküls in Lösung größer ist als $|\boldsymbol{\mu}_0|$. Der experimentell geschätzte Mittelwert von $(2{,}9 \pm 0{,}6)\,\mathrm{D}$ [52] hat allerdings einen großen Fehler, und ist nicht mit der experimentell gut bestimmten [43] Dielektrizitätskonstante $\epsilon_{\mathrm{H_2O}}$ kompatibel [53]. QM/MM-Computersimulationen sagen um etwa 10% kleinere Mittelwerte des Gesamtdipols voraus [33, 50, 54, 55].

Durch Dipol-Dipol-Wechselwirkungen bildet sich in kondensierter Phase ein dynamisches Netzwerk intermolekularer Wasserstoffbrücken [9], dessen typische tetraedrische Struktur in Abbildung 1.1b für einen kleinen Wassercluster durch die gestrichelten Linien skizziert ist. Da die Stärke einer Wasserstoffbrücke bei Raumtemperatur etwa beim zehnfachen der thermischen Energie liegt, ist Wasser unter physiologischen Bedingungen flüssig [42]. Viele Anomalien des Wasser können durch die mikroskopischen Eigenschaften des Wasserstoffbrückennetzwerks erklärt werden [42, 56].

1.1.2 Solvatisierungseffekte auf Amidgruppen

Wie im einleitenden Text dieses Abschnitts bereits erwähnt wurde, bildet sich das Proteinrückgrat durch die Synthese von Aminosäuren. Der linke Teil von Abbildung 1.2 zeigt zwei Aminosäuren AS1 und AS2, deren chemische Struktur bis auf die sog. Restgruppen (R1 bzw. R2) identisch ist: eine Aminogruppe ($\mathrm{NH_3^+}$), eine Carboxygruppe ($\mathrm{COO^-}$), und ein einzelnes H-Atom sind an ein zentrales Kohlenstoffatom (C_α) gebunden. Die Restgruppe definiert die Aminosäure sowie deren chemische und elektrostatische Eigenschaften [2], sowie deren Löslichkeit in Wasser [57].

Abbildung 1.2 illustriert weiterhin den Vorgang der *Peptidsynthese*, bei der die Carboxygruppe von AS1 und die Aminogruppe von AS2 unter Abspaltung eines $\mathrm{H_2O}$-Moleküls eine sog. Peptidbindung eingehen. Das rechts gezeigte resultierende Motiv, das sich durch die hervorgehobene *Amidgruppe* (AG) auszeichnet, kann nun mit weiteren Aminosäuren reagieren, was schließlich zur Ausbildung des Proteinrückgrats führt. Die AG ist planar und weist ein großes statisches Dipolmoment (Pfeil) auf.

Der Grund für diese charakteristischen Eigenschaften der AG liegt im über deren Atome delokalisierten π-Elektronensystem, das, wie in Abbildung 1.3a am Beispiel des

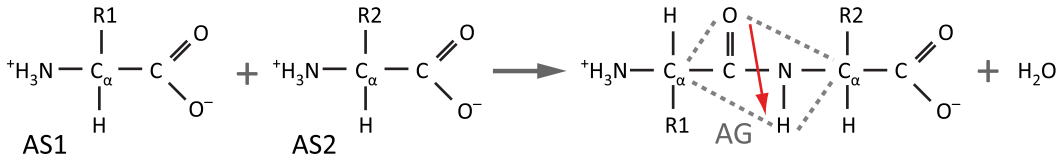


Abbildung 1.2: Zwei Aminosäuren AS1 und AS2 gehen unter Wasserabspaltung eine Peptidbindung ein, die resultierende planare AG (grau gestrichelt), deren starkes Dipolmoment durch den roten Pfeil symbolisiert ist, ist die dem Proteinrückgrat zugrundeliegende Struktur. In Wasser liegen die Amino- und Carboxygruppen überwiegend in der zwitterionischen Struktur vor.

molekularen Minimalmodells einer AG, dem N-Methyl-Acetamid (NMA), skizziert ist, zwei Resonanzstrukturen aufweist [2]. In der neutralen Resonanzstruktur A liegt die C′–N-Bindung als Einfachbindung und die C′=O-Bindung als Doppelbindung vor, in der zwitterionischen dipolaren Resonanzstruktur B hat hingegen die C′=O-Bindung Einfachbindungs- und die C′–N-Bindung Doppelbindungscharakter. Diese dort vorliegende torsionsstabile C′=N-Doppelbindung hält die AG planar, weswegen diese auch als *Peptidplättchen* bezeichnet wird. Durch äußere Felder kann die Auftretenswahrscheinlichkeit der dipolaren Resonanzstruktur B, und damit das Nettodipolmoment der AG erhöht werden: AGn sind also, ebenso wie Wassermoleküle, stark polarisierbar.

Abbildung 1.3b zeigt eine typische molekulare Konfiguration von in Wasser gelöstem NMA. In wässriger Lösung sind im Mittel am O-Atom der AG etwas mehr als zwei Wassermoleküle, und am H-Atom der AG ein Wassermolekül über Wasserstoffbrücken gebunden, wie eine Analyse der mit DFT-MD berechneten radialen Verteilung von flüssigem Wasser um ein NMA-Molekül ergab [58]. Die resultierende strukturierte Wasserumgebung generiert starke lokale elektrische Felder, die die AG polarisieren. Die aus Abbildung 1.3a qualitativ abgeleitete Polarisierbarkeit der AG wird durch QM-Rechnungen bestätigt: diese sagen beim Transfer von NMA aus dem Vakuum in wässrige Lösung eine Erhöhung des Dipolmoments um bis zu 75% voraus [59, 60]. Das mittlere Dipolmoment eines Wassermoleküls wächst dagegen beim Transfer in die flüssige Phase nur um etwa 40% (vgl. Abschnitt 1.1.1) [33, 50, 54, 55].

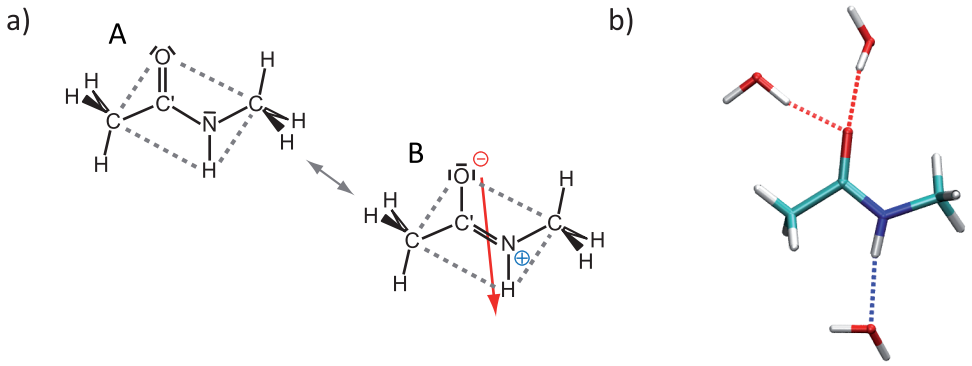


Abbildung 1.3: a) Neutrale (A) und dipolare (B) elektronische Resonanzstruktur der durch grau gestrichelte Linien angedeuteten AG von NMA. b) Das NMA und drei über Wasserstoffbrücken gebundene Wassermoleküle. Kohlenstoffatome (C) sind in cyan, Sauerstoffatome (O) in rot, Stickstoffatome (N) in blau und Wasserstoffatome (H) in weiß dargestellt.

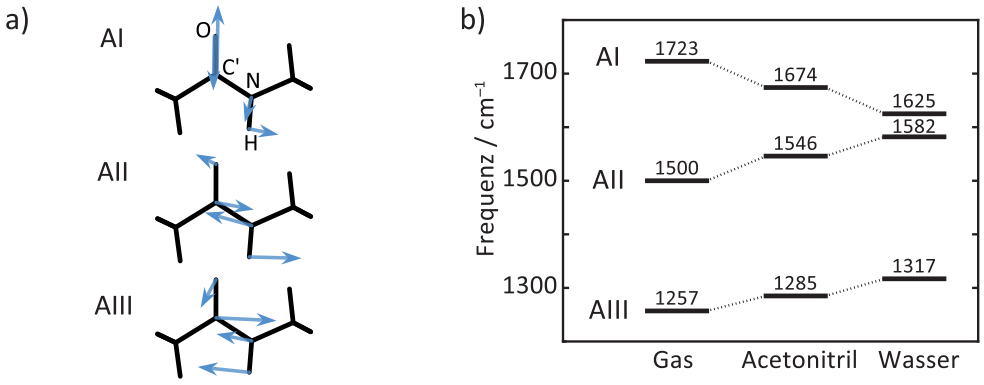


Abbildung 1.4: a) Für die Amidmoden AI-AIII von NMA charakteristische relative Atombewegungen innerhalb der AG. b) Experimentell beobachtete [10] solvatochrome Verschiebungen der Moden AI-AIII beim Transfer von NMA aus der Gasphase in Lösungsmittel zunehmender Polarität.

Die Polarisation einer AG durch äußere Felder hat, neben der Erhöhung ihres Dipolmoments, auch erheblichen Einfluß auf die intramolekularen Kräfte und damit auf ihre spektroskopischen Eigenschaften. Abbildung 1.4a illustriert am Beispiel von NMA drei charakteristische IR-aktive Normalmoden [61] einer AG, die als die *Amidmoden* AI, AII und AIII bezeichnet werden. Hierbei skizzieren die Pfeile die einer Normalmode zugrundeliegenden Atombewegungen, so besteht z.B. die AI-Mode hauptsächlich aus einer C′=O-Streckschwingung. Abbildung 1.4b zeigt die für NMA in Umgebungen unterschiedlicher Polarität experimentell beobachteten Frequenzen dieser Moden AI-AIII [10]. Beim Transfer des NMA von der Gasphase in das polare Lösungsmittel Acetonitril (CH_3CN, Dielektrizitätskonstante $\epsilon_{CH_3CN} \approx 36$ [62]) und in das stärker polare H_2O ($\epsilon_{H_2O} \approx 2\,\epsilon_{CH_3CN}$) verschieben sich die Moden AI bzw. AII und AIII stark und in unterschiedliche Richtungen.

Auch dieser *solvatochrome Effekt* kann anhand des Resonanzstrukturmodells der AG (siehe Abb. 1.3a) erklärt werden [63]. Mit steigender Lösungsmittelpolarität erhöht sich die Wahrscheinlichkeit für das Auftreten der polaren Resonanzstruktur B gegenüber der neutralen Struktur A. Die C′=O-Bindung hat also vermehrt Einfachbindungs-, die C′–N-Bindung vermehrt Doppelbindungscharakter. Damit nimmt die Stärke der C′=O-Kraftkonstante im Mittel ab, was eine Absenkung der Frequenz der AI-Mode zur Folge hat. Gleichzeitig nimmt die Stärke der C′–N-Bindung zu, und die Frequenzen der AII- und AIII-Moden, an denen die C′–N-Streckschwingung großen Anteil hat (vgl. Abb. 1.4a), steigen.

Das Schwingungsspektrum einer AG kann also Rückschlüsse auf deren Umgebung liefern, die in einem Polypeptid einerseits durch die jeweilige Lösungsmittelumgebung, andererseits durch die anderen AGn (und deren Restgruppen) gegeben ist.

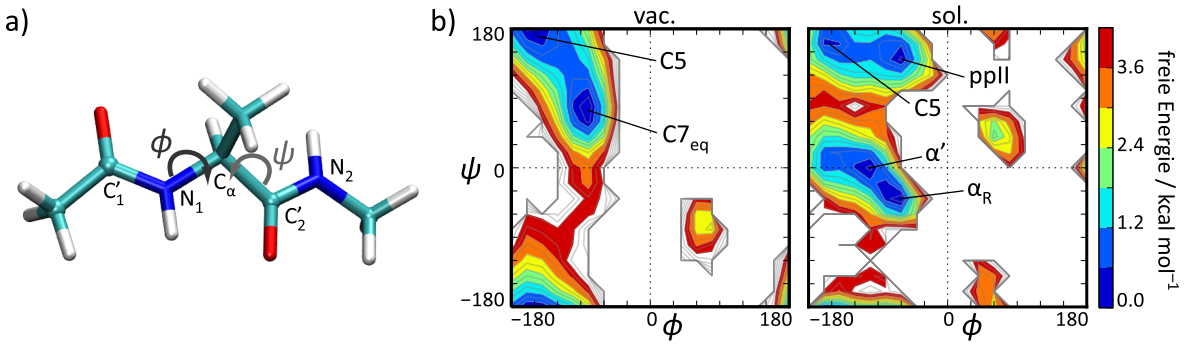


Abbildung 1.5: a) Alanindipeptid in der *C5*-Konformation, die durch die Diederwinkel $(\phi, \psi) \approx (-150°, +180°)$ definiert ist. b) Ramachandranplots der freien Energie $F(\phi, \psi)$ von Alanindipeptid im Vakuum (vac.) und in wässriger Lösung (sol.) bei einer Temperatur von 300 K. Weiße Bereiche kennzeichnen Regionen mit $F > 4{,}2$ kcal/mol (Simulationsmodelle: CHARMM22 bzw. CHARMM22+CMAP/TIP3P [66–68]).

1.1.3 Solvatisierung von Di- und Polypeptiden

Die eben diskutierten AGn treten nicht nur mit dem umgebenden Lösungsmittel, sondern auch untereinander über ihre Dipolmomente in Wechselwirkung, was die für die Ausbildung der Sekundärstrukturen grundlegende Wechselwirkung ist [64, 65].

Abbildung 1.5a zeigt ein einfaches Modellbeispiel für den Aufbau der Rückgratstruktur eines Polypeptids durch Verkettung von Aminosäuren (vgl. Abb. 1.2) über ein zentrales C_α-Atom: das Molekül Alanindipeptid (Ace-Ala-NHMe). Hier sind zwei AGn über ein C_α-Atom verbunden, an das außerdem eine Methyl-Restgruppe und ein H-Atom gebunden ist. Jeweils eine Methylgruppe sättigt den N- bzw. den C-Terminus ab. Im Gegensatz zu den torsionsstabilen Doppelbindungen innerhalb der AGn sind Rotationen um die N–C_α- bzw. C_α–C′-Einfachbindungen relativ leicht möglich, die Rückgratstruktur lässt sich daher durch die skizzierten *Diederwinkel* ϕ und ψ charakterisieren [2]. Der hochdimensionale Konfigurationsraum von Alanindipeptid wird so auf einen zweidimensionalen *Konformationsraum* reduziert.

Abbildung 1.5b zeigt exemplarisch zwei mittels MM-MD-Simulationen im Vakuum (vac.) und in wässriger Lösung (sol.) für Alanindipeptid bei Raumtemperatur generierte sog. Ramachandranplots [69] der freien Energie $F(\phi, \psi) \sim -\ln[\, p(\phi, \psi)\,]$, die aus der Verteilung $p(\phi, \psi)$ der Diederwinkel berechnet werden kann. Die mit in der Literatur üblichen Bezeichnungen [70] markierten lokalen Minima von F sind somit die wahrscheinlichsten Zustände im Konformationsraum.

Ein Vergleich der beiden Ramachandranplots zeigt den Einfluß wässriger Lösung auf die Konformationslandschaft von Alanindipeptid. Im Vakuum herrschen die in Abbildung 1.5a gezeigte *C5*- und die *C7*$_{eq}$-Konformation vor, in denen die Dipolmomente der beiden AGn energetisch günstig antiparallel zueinander angeordnet sind. Parallele Anordnungen der Dipole, wie z.B. im α-Bereich, kommen im Vakuum äußerst selten vor. In Wasser wechselwirken die Dipolmomente der AGn zusätzlich mit den umgebenden Wassermolekülen, auf diese Weise wird die freie Energie einiger im Vakuum ungünstiger

Motive abgesenkt. Wasserstoffbrückenbindungen innerhalb des Dipeptids, die teilweise durch Wassermoleküle unterstützt werden, führen zur Ausdifferenzierung weiterer Strukturen wie *ppII*, die zusammen mit *C5* den sog. β-Bereich bilden.

Die lokalen Minima von $F(\phi, \psi)$ sind durch teilweise hohe Energiebarrieren voneinander getrennt [71], so sind z.B. spontane Übergänge von Regionen mit $\phi < 0°$ zu Regionen mit $\phi > 0°$ bei Raumtemperatur sehr selten. Viele (ϕ,ψ)-Kombinationen sind durch sterische Hinderungen, d.h. Pauli-Abstoßungen der atomaren Elektronendichten, verboten [2]. Sterische Hinderungen sind auch der Grund, warum die Restgruppen der AGn Einfluß auf die Beweglichkeit des Rückgrats und auf die Konformationslandschaft haben [70].

In größeren Polypeptiden interagieren nicht nur benachbarte AGn, sondern, ermöglicht durch Krümmungen des Rückgrats, auch weiter entfernte [2]. Die wichtigsten Sekundärstrukturmotive in Proteinen sind die α-Helix, bei der die parallel angeordneten Dipole der AGn einen Makrodipol bilden, sowie das β-Faltblatt, das kein Nettodipolmoment aufweist [64, 65]. Zur Strukturbildung können außerdem Wechselwirkungen zwischen Restgruppen beitragen [2].

Die strukturtypischen, von anderen AGn bzw. Wassermolekülen erzeugten lokalen elektrischen Felder führen zu spezifischen Polarisationen der einzelnen AGn, die sich, wie oben diskutiert, auf deren Schwingungsmuster auswirken [10]. So sind die verschiedenen Sekundärstrukturen durch (IR-)Spektroskopie [11] der Amidmoden (siehe Abb. 1.4a) unterscheidbar [26, 72, 73], mit zeitaufgelösten Messungen kann die Konformationsdynamik untersucht werden [19, 21, 74–76]. Die Interpretation der experimentellen Beobachtungen, z.B. der Schluss von einem gemessenen Schwingungsspektrum auf eine bestimmte Proteinstruktur, kann durch theoretische Methoden unterstützt werden, die nun im Folgenden vorgestellt werden.

1.2 Molekulardynamik-Simulationen

Einen wichtigen theoretischen Zugang zu Biomolekülen stellen die im Jahre 1957 durch Alder und Wainwright [12] erstmals angewendeten MD-Simulationen dar [13–15]. Mittels einer solchen Simulation wird eine zeitdiskrete Trajektorie $\mathbf{R}(t) = [\mathbf{r}_1(t), \ldots, \mathbf{r}_N(t)]^{\mathrm{T}}$ eines Systems aus N Atomen generiert, die zu jedem Zeitpunkt t alle Atomkoordinaten $\mathbf{r}_i \in \mathbb{R}^3$ $(i = 1, \ldots, N)$ speichert, und die dann der jeweiligen Fragestellung entsprechend analysiert werden kann [14].

Die sog. Born-Oppenheimer-MD [77] stützt sich auf die gleichnamige adiabatische Näherung [78], die eine Trennung der Zeitskalen, auf denen die Bewegungen der Atomkerne und der viel leichteren Elektronen stattfinden, ermöglicht. Die Atome i werden dann als Punkte der Masse m_i beschrieben, die sich gemäß der klassischen Newton'schen Mechanik bewegen [79], d.h. es gilt

$$m_i \ddot{\mathbf{r}}_i = -\nabla_i U(\mathbf{R}). \tag{1.1}$$

Die auf die Atome i wirkenden Kräfte auf der rechten Seite von Gleichung (1.1) werden dabei als Gradienten einer noch zu spezifizierenden potentiellen Energie $U(\mathbf{R})$ berechnet.

Bei einer numerischen Integration von Gleichung (1.1), z.B. mittels des Verlet-Algorithmus [80, 81], muss der verwendete Zeitschritt Δt klein genug gewählt werden, um auch die schnellsten Freiheitsgrade im System glatt abzutasten [14]. Da dieser typischerweise in der Größenordnung von Femtosekunden ($10^{-15}\,\mathrm{s}$) liegt, sind zur Berechnung einer Trajektorie mit einer Dauer von einer Mikrosekunde bereits etwa eine Milliarde Integrationsschritte [und damit Auswertungen der rechten Seite von Glg. (1.1)] nötig. Für ein gegebenes N-Teilchensystem hängt die Realisierbarkeit einer solchen Trajektorie entscheidend vom für die Auswertung von $U(\mathbf{R})$ benötigten Rechenaufwand ab.

Das für MD-Simulationen nötige Potential $U(\mathbf{R})$ steht nun im Fokus der weiteren Arbeit. Da eine analytische quantenmechanische Formulierung dieses Potentials für mehratomige Systeme meist unmöglich ist [82], muss man für seine Beschreibung auf geeignete Approximationen zurückgreifen.

1.2.1 Molekülmechanische Kraftfelder

Eine vergleichsweise grobe Näherung der physikalischen Realität stellen die MM-Kraftfelder dar [83, 84]. Deren Grundidee ist es, das in (1.1) eingehende Potential

$$U^{\mathrm{MM}}(\mathbf{R}) = U_{\mathrm{b}}(\mathbf{R}) + U_{\mathrm{nb}}(\mathbf{R}) \tag{1.2}$$

in lokale Wechselwirkungen U_{b} naher chemisch gebundener Atome und in langreichweitige Wechselwirkungen U_{nb} weiter entfernter bzw. chemisch nicht gebundener Atome aufzuteilen und die einzelnen Beiträge mit einfachen analytischen Formeln auszudrücken. Die erforderlichen Parameter werden, z.B. durch QM-Referenzrechnungen oder Vergleich mit experimentellen Daten empirisch bestimmt [83], und definieren dann ein (Protein-)Kraftfeld wie z.B. CHARMM22 [67]. Für Wasser existiert eine Vielzahl von speziellen MM-Kraftfeldern [66, 85, 86].

In U_{b} werden kovalente Bindungen und Winkelpotentiale z.B. durch harmonische Potentiale, und Torsionsbarrieren durch periodische Diederpotentiale approximiert. Die jeweiligen Parameter sind in konventionellen MM-Kraftfeldern konstant. Der langreichweitige Anteil

$$U_{\mathrm{nb}}(\mathbf{R}) = U_{\mathrm{vdW}}(\mathbf{R}) + U_{\mathrm{elec}}(\mathbf{R}) \tag{1.3}$$

setzt sich zusammen aus den van der Waals-Wechselwirkungen U_{vdW} und den elektrostatischen Wechselwirkungen U_{elec}.

Zur Beschreibung der van der Waals-Wechselwirkungen U_{vdW} zwischen sich im Abstand $r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$ befindenden Atomen i und j werden häufig empirische 12-6 Lennard-Jones-Potentiale [87] $U_{\mathrm{LJ}}(\mathbf{R}) = \sum_{i<j} \left[A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6}\right]$, in selteneren Fällen Buckingham-Potentiale [88] $U_{\mathrm{Bu}}(\mathbf{R}) = \sum_{i<j} \left[A_{ij}^{1} \exp(-A_{ij}^{2} r_{ij}) - B_{ij}/r_{ij}^{6}\right]$ gewählt. Die Parameter A_{ij} bzw. (A_{ij}^{1}, A_{ij}^{2}) beschreiben hier die Pauli-Abstoßung zwischen den Atomen i und j, während die interatomare Dispersionsattraktion durch die Parameter B_{ij} modelliert wird. Alle Summationen über Atompaare schließen üblicherweise langreichweitige Wechselwirkungen zwischen gebundenen Atomen innerhalb etwa zwei kovalenter Bindungen aus (sog. 1-M-exclusion, $M = 2, 3, \ldots$) [67].

Die elektrostatischen Wechselwirkungen U_{elec} werden in gängigen MM-Kraftfeldern durch punktförmige Partialladungen q_i an den Atomorten $\mathbf{r}_i$ modelliert [67, 83, 86]. Deren Energiebeitrag ist durch die Coulombsumme

$$U_{\text{elec}}^{\text{MM}}(\mathbf{R}) = \frac{1}{2} \sum_{i,j \neq i} q_i \Phi(\mathbf{r}_i \,|\, q_j, \mathbf{r}_j) \tag{1.4}$$

gegeben, in der

$$\Phi(\mathbf{r} \,|\, q_j, \mathbf{r}_j) = \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \tag{1.5}$$

das durch die sich an der Position $\mathbf{r}_j$ befindende Punktladung q_j am Ort $\mathbf{r}$ erzeugte elektrostatische Potential bezeichnet [89]. Punktladungen sind recheneffizient behandelbar, stellen aber eine grobe Approximation für die ausgedehnten atomaren Ladungsverteilungen dar. Einige spezielle Kraftfelder, vor allem für Wasser [90, 91], verwenden daher gaußförmig verbreiterte Ladungen, für die ein analoger Ausdruck gilt [92].

Die Modellierung der elektrostatischen Wechselwirkungen ausschließlich durch statische Ladungen kann Polarisationseffekte auf die Ladungsverteilung allerdings nur im Mittel, d.h. im Rahmen einer Molekularfeld-Näherung (engl. *mean-field*) berücksichtigen [93]. Dies gilt gleichermaßen für die Verwendung feldunabhängiger Potentialparameter in gebundenen Wechselwirkungen. So können die elektrostatischen und spektroskopischen Eigenschaften vor allem stark polarisierbarer Verbindungen wie Wasser (siehe Abschnitt 1.1.1) oder AGn (siehe Abschnitt 1.1.2) nur unzureichend beschrieben werden [29, 55, 63, 94].

Im speziellen Fall von nicht-polarisierbaren Wassermodellen schränkt die mean-field-Näherung deren Transferierbarkeit, z.B. in die Umgebung lokal stark polarer oder geladener Proteine, und damit deren Verwendbarkeit als akkurate Lösungsmittelmodelle ein. Für die Beschreibung von Wasser gibt bereits einige polarisierbare Ansätze [55, 85, 90, 91, 95, 96]. Auch für die Modellierung von Polarisationseffekten auf gebundene Wechselwirkungen existieren Vorschläge [63, 94, 97–100]. Die Weiterentwicklung eines solchen *spektroskopischen Kraftfelds* [63, 94] ist Thema der laufenden Dissertation meines Kollegen Christoph Wichmann.

Polarisierbare Molekülmechanische Kraftfelder

Elektronische Polarisationseffekte auf eine molekulare Ladungsverteilung können mittels verschiedener Ansätze explizit in einem Kraftfeld berücksichtigt werden [93, 101–103], z.B. durch Einführen von Ladungen variabler Stärke (*fluctuating-charge-Modelle* [104]), variabler Position (*Drude-Oszillatoren* [105]) oder von induzierbaren Dipolen [31, 106]. Die potentielle Energie

$$U^{\text{PMM}}(\mathbf{R}) = U^{\text{MM}}(\mathbf{R}) + U_{\text{elec}}^{\text{pol}}(\mathbf{R}) \tag{1.6}$$

eines solchen PMM-Kraftfelds beinhaltet dann die zusätzlichen elektrostatischen Wechselwirkungen $U_{\text{elec}}^{\text{pol}}$ [93, 101].

Durch die Einführung induzierbarer Dipole [106] an den Atomorten $\mathbf{r}_i$, deren Stärken

$$\mathbf{p}_i = \alpha_i \, \mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i) \tag{1.7}$$

mit den skalaren atomaren Polarisierbarkeiten α_i aus der linearen Antwortnäherung folgen, ergibt sich der zusätzliche Energiebeitrag [31]

$$U_{\text{elec}}^{\text{pol}}(\mathbf{R}) = \frac{1}{2} \sum_{i,j \neq i} q_i \Phi(\mathbf{r}_i \,|\, \mathbf{p}_j, \mathbf{r}_j) - \frac{1}{2} \sum_i \mathbf{p}_i \cdot \mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i) + \frac{1}{2} \sum_i \mathbf{p}_i^2 / \alpha_i. \tag{1.8}$$

Hier bezeichnet $\Phi(\mathbf{r} \,|\, \mathbf{p}_j, \mathbf{r}_j)$ das vom Dipol $\mathbf{p}_j$ am Ort $\mathbf{r}$ generierte Potential [89], und der Ausdruck

$$\mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i) = \sum_{j \neq i} \left[\mathbf{E}(\mathbf{r}_i \,|\, q_j, \mathbf{r}_j) + \mathbf{E}(\mathbf{r}_i \,|\, \mathbf{p}_j, \mathbf{r}_j) \right] \tag{1.9}$$

fasst die von den statischen Ladungen q_j und den induzierbaren Dipolen $\mathbf{p}_j$ $(j \neq i)$ am Ort $\mathbf{r}_i$ erzeugten elektrischen Felder $\mathbf{E}(\mathbf{r}_i \,|\, q_j, \mathbf{r}_j)$ bzw. $\mathbf{E}(\mathbf{r}_i \,|\, \mathbf{p}_j, \mathbf{r}_j)$ zusammen. Der letzte Term in Gleichung (1.8) ist die zur Erzeugung der Dipole $\mathbf{p}_i$ nötige sog. Selbstenergie.

Da das nach Gleichung (1.7) einen Dipol $\mathbf{p}_i$ induzierende Feld $\mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i)$ von allen anderen Dipolen $\mathbf{p}_j$ $(j \neq i)$ erzeugt wird, muss zur Berechnung der atomaren Kräfte eine selbstkonsistente Lösung gefunden werden (z.B. durch Iteration) [107, 108]. Dieser Umstand, sowie die zusätzlichen elektrostatischen Quellen sind für den erhöhten Rechenaufwand von PMM-Kraftfeldern gegenüber MM-Kraftfeldern verantwortlich [109]. Eine unphysikalische Überpolarisation wird durch die Verwendung im Nahfeld abgeschirmter induzierbarer Dipole vermieden [110]. Eine solche Abschirmung kann effektiv auch durch die Verwendung gaußförmig verbreiterter Dipole [55, 90, 111] erreicht werden, entsprechend gelten statt (1.8) und (1.9) modifizierte Formeln [112].

Im Fall von Wasser ist die explizite Berücksichtigung der Polarisierbarkeit alleine aber nicht ausreichend, um die wesentlichen Eigenschaften dieses Lösungsmittels zu beschreiben. Auch die höheren elektrostatischen Momente wie das Quadrupolmoment müssen, z.B. durch die Stärken und Positionen von Partialladungen, die nicht notwendigerweise an den Atomorten lokalisiert sind [66], korrekt modelliert werden [56, 113]. Sog. Vierpunktmodelle verschieben die negative Partialladung vom O-Atom auf die φ_{HOH}-Winkelhalbierende, Fünfpunktemodelle verteilen sie auf zwei senkrecht zur Molekülebene stehenden *lone-pair*-Positionen. Die mittels einer eleganten Strategie entwickelten polarisierbaren Vier- bzw. Fünfpunktemodelle TL4P und TL5P [55], nach der durch das Einbeziehen möglichst vieler experimentell gesicherter bzw. mit DFT berechneter mikroskopischer Eigenschaften des Wassermoleküls nur etwa ein Drittel der benötigten Parameter über empirische Optimierung makroskopischer Flüssigphaseneigenschaften bestimmt wurden, waren aber nicht in der Lage, die Dichteanomalie vorherzusagen [114]. Dieses Ergebnis ist ein Symptom für das von Wassermodellen solcher Geometrie nur unzureichend beschriebene Quadrupolmoment [56, 113], für dessen korrekte Modellierung mindestens die durch eine Kombination aus Vier- und Fünfpunktemodellgeometrie (also ein Sechspunktmodell) zur Verfügung stehenden fünf Partialladungen benötigt werden.

Effiziente Behandlung Langreichweitiger Wechselwirkungen

Neben der Entwicklung einer geeigneten Strategie zur Festlegung der zahlreichen freien Parameter in einem (P)MM-Kraftfeld stellen vor allem die langreichweitigen Wechselwirkungen (1.3) eine Herausforderung für theoretische Beschreibungen dar [14, 84, 115]. Der Rechenaufwand für die exakte Auswertung von Paarwechselwirkungen [z.B. (1.4), (1.8), oder der Modellpotentiale für van der Waals-Wechselwirkungen] skaliert mit $\mathcal{O}(N^2)$, was die Verwendung von Näherungsmethoden nötig macht.

Ein einfaches Abschneiden dieser Wechselwirkungen führt im Falle der Elektrostatik zu großen systematischen Fehlern, da das durch das Lösungsmittel ausgeübte Reaktionsfeld dann nicht vollständig beschrieben ist [116]. Das Abschneiden der van der Waals-Wechselwirkungen zieht algorithmische Heiz- oder Abkühl-Artefakte nach sich [109]. Ansätze wie die Ewald-Summation [117–119] oder schnelle Multipolmethoden [120–122] (engl. *fast multipole method*, FMM) vermeiden solche Artefakte und ermöglichen eine mit $\mathcal{O}(N \log N)$ bzw. mit $\mathcal{O}(N)$ skalierende approximative Berechnung von Paarwechselwirkungen.

Die in dieser Arbeit verwendete *schnelle strukturadaptierte Multipolmethode* (SAMM) [109, 116, 123–127] erlaubt eine strikt linear skalierende Berechnung [109] aller in den oben beschriebenen (P)MM-Kraftfeldern auftretenden Paarwechselwirkungen. Durch die Verwendung zweiseitiger Taylorentwicklungen der Ordnungen p, q, bzw. r für Elektrostatik-, Dispersions- und Lennard-Jones-Repulsionswechselwirkungen können die auftretenden Kräfte approximativ mit kontrollierter Genauigkeit als exakte Gradienten der entsprechenden Energien berechnet werden [127] (Standardwerte: $p = 4$, $q = 3$, $r = 1$). SAMM erlaubt in nicht-periodischen Systemen eine Energie- und drehimpulserhaltende Dynamik [127].

Da der SAMM-Algorithmus in dieser Arbeit zur effizienten Berechnung von DFT/PMM-Wechselwirkungen eingesetzt wird, werden im Folgenden dessen Grundkonzepte anhand der in einem MM-Kraftfeld auftretenden elektrostatischen Paarwechselwirkungen (1.4) skizziert.

Der SAMM-Algorithmus

Abbildung 1.6 zeigt einen Ausschnitt aus einer Gruppe von MM-Wassermolekülen, deren gegebene Ladungsverteilung (Punkte) durch SAMM in ein hierarchisches geschachteltes Netzwerk aus Ladungsclustern (gestrichelte Kreise, c_1^0-c_6^0, c_1^1-c_3^1) zerlegt wurde. Die Ausdehnung dieser Cluster wächst mit zunehmender Hierarchiestufe $l = 0, 1, \ldots, l^{\text{top}}$ bis zu einer vordefinierten Maximalstufe l^{top}. Möglichst kompakte Cluster werden durch den Einsatz neuronaler Verfahren [128] und durch die periodische Aktualisierung der Clusterstruktur erreicht [109]. Auf der untersten Hierarchiestufe $l = 0$ befinden sich in SAMM vordefinierte chemische Gruppen [123], die sog. strukturellen Einheiten, die im betrachteten Fall durch die einzelnen Wassermoleküle gegeben sind.

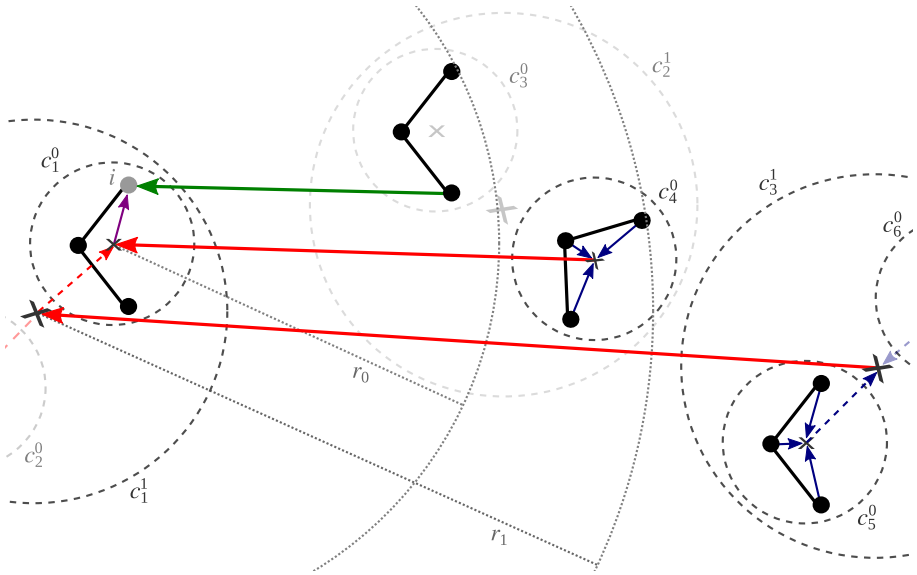


Abbildung 1.6: Approximative Berechnung des durch die atomaren Ladungen (schwarze Punkte) von umgebenden Wassermolekülen erzeugten Potentials am Ort des Atoms i (grauer Punkt) mit SAMM. Gestrichelte Kreise definieren Ladungscluster c_I^l, die gepunkteten Kreissegmente (Radien r_0 bzw. r_1) markieren die durch das IAC gegebene Grenzen zwischen den SAMM-Hierarchiestufen. Pfeile skizzieren die Berechnung eines exakten Coulombpotentials (grün), bzw. von Multipolmomenten (blau), von Taylorentwicklungskoeffizienten (rot), sowie die Auswertung einer lokalen Taylorentwicklung (lila).

Die Entscheidung, ob die Wechselwirkung zweier Cluster c_I^l und $c_{I'}^l$ auf einer bestimmten Hierarchiestufe l erlaubt ist, erfolgt durch Prüfung des Akzeptanzkriteriums [109]

$$\frac{1}{2}\left[\vartheta_{I,l}(r) + \vartheta_{I',l}(r)\right] \leq \Theta_\chi, \tag{1.10}$$

dem sog. IAC (für *interaction acceptance criterion*). Dazu werden die genauigkeitsgewichteten apparenten Ausdehnungen $\vartheta_{I,l}(r) \equiv 2R_{I,l}/(r\, a_{I,l})$ der Cluster c_I^l und $c_{I'}^l$, die aus deren Gyradionsradien $R_{I,l}$, deren Abstand r, und Gewichtsparametern $a_{I,l}$ berechnet werden, mit einer vordefinierten Schwelle Θ_χ ($\chi \in \{\mathrm{f}, \mathrm{m}, \mathrm{a}\}$, s.u.) verglichen [109]. In die Gewichtsparameter $a_{I,l}$ geht die elektrostatische Signatur und die Geometrie des jeweiligen Clusters ein, was die Genauigkeit der Berechnung nahezu unabhängig von der chemischen Zusammensetzung des Systems macht [109]. Das Prüfen des IAC (1.10) beginnt auf der höchsten Stufe l^{top}. Ist eine Wechselwirkung auf einer Stufe l nicht erlaubt, werden die Cluster in ihre Bestandteile zerlegt, und das IAC dafür auf Stufe $l-1$ geprüft. Die in Abbildung 1.6 grau gepunkteten Radien r_1 und r_0 illustrieren die durch das IAC gegebenen Minimalabstände für Wechselwirkungen des Clusters c_1^1 auf Stufe $l = 1$ bzw. des Clusters c_0^1 auf Stufe $l = 0$.

Die Berechnung des durch die Ladungsverteilung am Ort des grau hervorgehobenen Atoms i generierten Potentials (und den zur Kraftberechnung nötigen Potentialgradienten) erfolgt dann auf allen Stufen $l \geq 0$ approximativ über Multipolentwicklungen (blaue Pfeile) und Taylorentwicklungen (rote und lila Pfeile) um die geometrischen Zentren „×" der Cluster. Die gestrichelten blauen und roten Pfeile bezeichnen das Verschieben von Multipolmomenten bzw. Taylorentwicklungskoeffizienten zwischen den Hierarchie-

stufen. Nur wenn eine approximative Cluster-Cluster-Wechselwirkung auch auf Stufe $l = 0$ durch das IAC (1.10) verboten ist, muss die exakte Auswertung des Coulombausdrucks (1.5) erfolgen (der grüne Pfeil illustriert eine solche Auswertung). Die genauen mathematischen Details finden sich in Ref. [126].

Im von Lorenzen *et al.* [109, 126, 127] weiterentwickelten SAMM$^{\chi}$-Algorithmus wird die Taylorentwicklung eines von einem Multipolmoment m'ter Ordnung generierten Potentials bis zur Ordnung $n = 4-m$ ausgeführt, da die so berechneten Kräfte dann das Newton'sche Reaktionsprinzip erfüllen [126]. Im Gegensatz zu fixen Distanzklassen [116, 125] ermöglicht das IAC (1.10) eine genaue Kontrolle der durch die FMM-Approximationen verursachten Fehler. Es stehen die drei Kompromisse Θ_{f} (hohe Effizienz), Θ_{m} (mittlere Effizienz, mittlere Genauigkeit) und Θ_{a} (hohe Genauigkeit) aus Genauigkeit und Recheneffizienz zur Verfügung [109]. Zum Beispiel erlaubt das IAC für flüssiges Wasser einen Übergang von exakten Ausdrücken zu den FMM-Approximationen bereits bei kleinen Atomabständen um $\approx 5\,\text{Å}$ was einen deutlichen Effizienzgewinn gegenüber früheren SAMM-Algorithmen [116, 123–125] bietet und so z.B. die effiziente Behandlung lokal sehr komplexer PMM-Mehrpunktmodelle für Wasser [55, 129] ermöglicht [109].

Die in Simulationssystemen endlicher Größe auftretenden Oberflächeneffekte werden bei der Verwendung von SAMM durch die Annahme toroidaler Randbedingungen ausgeschlossen [14]. Um Periodizitätsartefakte zu vermeiden, treten gemäß der Minimalbild-Konvention (engl. *minimum image convention*, MIC) nur Atome innerhalb eines Abstands r_{MIC}, der gerade der Hälfte des Durchmessers der verwendeten Einheitszelle entspricht, in Wechselwirkung [14]. Ein hartes Abschneiden der Wechselwirkungen wird durch die Einführung eines dielektrisches Kontinuum außerhalb von r_{MIC} vermieden [116], was z.B. im Falle von Wasser ab einer Entfernung von $15\,\text{Å}$ in sehr guter Näherung möglich ist [130].

Da in (P)MM-MD Simulationen von Molekülen in explizitem Lösungsmittel der Großteil der Rechenzeit trotzdem in die — oft uninteressante — Lösungsmitteldynamik investiert werden muss, kann die Recheneffizienz durch das Ersetzen der Lösungsmittelmoleküle durch ein Kontinuum erheblich gesteigert werden [92, 131]. Ein solches *implizites Lösungsmittel*, wie die kürzlich im Rahmen der Dissertation von Sebastian Bauer entwickelte energieerhaltende HADES-Methode (für *Hamiltonian dielectric solvent*), erweitert somit die durch Simulation zugänglichen Längen- und Zeitskalen, und erleichtert so z.B. Studien zu Faltungsprozessen [132]. Zur genauen Berechnung von Schwingungsspektren müssen allerdings explizite Lösungsmittelmodelle verwendet werden, da neben der Beschreibung der nicht-lokalen dielektrischen Eigenschaften auch lokale spezifische mikroskopische Effekte wie Wasserstoffbrückenbindungen erfassst werden müssen [133].

Anwendungsbereiche von (P)MM-Kraftfeldern

Die skizzierten einfachen (P)MM-Kraftfelder erlauben eine sehr effiziente Auswertung des die Dynamik (1.1) bestimmenden Potentials $U(\mathbf{R})$, und ermöglichen so die Beschreibung großer Biomoleküle in ausgedehnten Lösungsmittelumgebungen. So konnte z.B. mittels MM-MD-Simulation ein virtuelles Kraftmikroskopie-Experiment [11] an ei-

nem Komplex aus dem aus 159 Aminosäuren bestehenden Protein Streptavidin und dem Vitamin Biotin durchgeführt werden [18]. Die Simulation des in wässriger Lösung eingebettete Komplexes (ca. 11.000 Atome insgesamt) ermöglichte einen detaillierten Einblick in die Wechselwirkung der beiden Moleküle während des Auseinanderreißens. Eine ähnliche Studie wurde auch am größten bekannten im menschlichen Körper vorkommenden Protein, dem Titin, durchgeführt [134]. Auch eine komplette, wenn auch primitive, Lebensform, das Tabakmosaikvirus konnte bereits simuliert werden [135]. Die zugänglichen Systemgrößen und Zeitskalen werden durch algorithmische Fortschritte und wachsende Rechnerleistung weiter steigen [136].

Angesichts dieser Erfolge der MM-MD darf aber nicht außer Acht gelassen werden, dass die Qualität der Beschreibung entscheidend von der Wahl der Modellklasse (z.B. MM oder PMM) und -parameter abhängt [29, 83, 137]. Ferner können die harmonischen Bindungspotentiale chemische Reaktionen wie das Bilden oder Lösen kovalenter Bindungen nicht beschreiben. Wie schon erwähnt, sind Kraftfelder, die die Polarisationseffekte auf Bindungspotentiale vernachlässigen, für die Berechnung von Proteinschwingungsspektren ungeeignet [63, 94]. Die Verbesserung der MM-Kraftfelder ist daher immer noch ein aktives Forschungsgebiet [138], das sich dabei auf genaue quantenmechanische Beschreibungen stützen kann.

1.2.2 Dichtefunktionaltheorie

Eine quantenmechanische Berechnung des Potentials $U(\mathbf{R})$ in Gleichung (1.1) für eine gegebene atomare Kernkonfiguration $\mathbf{R}$ beruht im Rahmen der Born-Oppenheimer-Näherung [78] auf der (approximativen) Lösung der nicht-relativistischen zeitunabhängigen Schrödingergleichung

$$\hat{H}\Psi = E\Psi, \tag{1.11}$$

in der $E \equiv E(\mathbf{R})$ die elektronische Energie und Ψ die elektronische Wellenfunktion darstellt [139]. Der elektronische Hamiltonoperator $\hat{H} \equiv \hat{H}(\mathbf{R}) = \hat{T} + \hat{V}_{\mathrm{ke}}(\mathbf{R}) + \hat{V}_{\mathrm{ee}}$ setzt sich aus der kinetischen Energie $\hat{T}$ der Elektronen, sowie den Kern-Elektron- und Elektron-Elektron-Wechselwirkungen $\hat{V}_{\mathrm{ke}}$ bzw. $\hat{V}_{\mathrm{ee}}$ zusammen. Das effektive Potential $U^{\mathrm{QM}}(\mathbf{R}) = E(\mathbf{R}) + V_{\mathrm{kk}}(\mathbf{R})$, aus dem dann in (1.1) die Kräfte auf die Atomkerne abgeleitet werden können, erhält man durch Addition der elektronischen Energie und der Kern-Kern-Wechselwirkung V_{kk}. Die elektronische Wellenfunktion $\Psi \equiv \Psi(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ eines n-Elektronensystems hängt für ein gegebenes $\mathbf{R}$ von n elektronischen Orts- und Spinkoordinaten $\mathbf{x}_k \equiv (\mathbf{r}_k, s_k)$, $k = (1, \ldots, n)$, ab und muss Eigenschaften wie Antisymmetrie und Normierbarkeit aufweisen sowie die jeweiligen Randbedingungen des Systems erfüllen [139]. Es existieren eine Reihe von *ab initio* Verfahren [77, 140] zur approximativen Lösung der Schrödingergleichung (1.11).

Die Hohenberg-Kohn-Theoreme

Ist man nur an molekularen Eigenschaften im elektronischen Grundzustand interessiert, liefert die DFT einen guten Kompromiss aus Genauigkeit und Recheneffizienz. Das erste

der entsprechenden von Hohenberg und Kohn aufgestellten Theoreme besagt, dass die Elektronendichte $\rho(\mathbf{r})$ alle Grundzustandseigenschaften eines Moleküls definiert [141]. So können Observablen wie die kinetische Energie $T[\rho]$ der Elektronen, die Elektron-Elektron-Wechselwirkung $V_{\mathrm{ee}}[\rho]$ oder die elektronische Gesamtenergie

$$E[\rho] = \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} + F_{\mathrm{HK}}[\rho] \tag{1.12}$$

als Funktionale der dreidimensionalen Größe ρ ausgedrückt werden, ohne eine hochdimensionale komplexe Wellenfunktion Ψ als Lösung von (1.11) bestimmen zu müssen. Das elektrostatische Potential $v(\mathbf{r})$ im ersten Term von Gleichung (1.12) wird durch die atomaren Kernladungen erzeugt, und kann außerdem weitere Beiträge $\Phi(\mathbf{r})$ durch äußere Potentiale beinhalten (liegen keine äußeren Felder an, entspricht dieser Term der Kern-Elektron-Wechselwirkung $V_{\mathrm{ke}}[\rho]$). Für die beiden Beiträge zum *Hohenberg-Kohn-Funktional* $F_{\mathrm{HK}}[\rho] = T[\rho] + V_{\mathrm{ee}}[\rho]$, sind allerdings keine exakten analytische Formen bekannt. Dieses unbekannte Funktional, das Becke in einem sehr lesenswerten Übersichtsartikel [142] als „heiligen Gral der Elektronenstrukturtheorie" bezeichnet, ist universell, enthält also keine molekülspezifischen Annahmen.

Das zweite Hohenberg-Kohn-Theorem besagt [141], dass jede nicht-negative, normierte ($\int \bar{\rho}(\mathbf{r}) d\mathbf{r} = n$) Elektronendichte $\bar{\rho}$ das Variationsprinzip

$$E_0 \leq E[\bar{\rho}] \tag{1.13}$$

erfüllt. Die Grundzustandselektronendichte ρ_0 ist also diejenige, die die minimale (Grundzustands-)Energie $E_0 \equiv E[\rho_0]$ liefert. Eine detaillierte Einführung in die hier nur knapp skizzierten Grundideen der DFT findet sich in Ref. [139].

Das Kohn-Sham-Verfahren

Aufbauend auf den Hohenberg-Kohn-Theoremen entwickelten Kohn und Sham ein praktisch anwendbares Rechenverfahren [143]. Die Dichte ρ der Elektronen und eine Näherung $T_s[\rho]$ für deren kinetische Energie wird hier unter der Annahme nicht-wechselwirkender Elektronen über Einelektron-Spinorbitale ψ_k berechnet. Entsprechend formulierten sie das Hohenberg-Kohn-Funktional als

$$F_{\mathrm{HK}}[\rho] = T_s[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho], \tag{1.14}$$

mit der klassischen Wechselwirkungsenergie $J[\rho]$ der Elektronen und dem Austausch-Korrelations-Funktional $E_{\mathrm{xc}}[\rho] = T[\rho] - T_s[\rho] + V_{\mathrm{ee}}[\rho] - J[\rho]$, das die Fehler der gemachten Approximationen sammelt.

Ist ein Ausdruck für E_{xc} gegeben (s.u.), so ergeben sich die Einelektron-Orbitale ψ_k aus der Lösung der n gekoppelten Eigenwertprobleme

$$\left[-\tfrac{1}{2}\nabla^2 + v_{\mathrm{eff}}(\mathbf{r})\right] \psi_k = \epsilon_k \psi_k, \tag{1.15}$$

in denen ϵ_k die Energieeigenwerte und $v_{\text{eff}}(\mathbf{r}) = v(\mathbf{r}) + \delta J[\rho]/\delta\rho + \delta E_{\text{xc}}[\rho]/\delta\rho$ das effektive Kohn-Sham-Potential bezeichnen. Ausgehend von einer initialen Annahme $\rho^{(1)}$ für die Elektronendichte wird v_{eff} berechnet und dann die ψ_k als Lösung des Gleichungssystems (1.15) bestimmt. Aus den ψ_k kann dann eine neue Elektronendichte $\rho^{(2)}$ und mit (1.12) und (1.14) die Gesamtenergie $E[\rho^{(2)}]$ berechnet werden. Dieser Prozess wird solange iteriert, bis die Änderung der Gesamtenergie einen kleinen Schwellwert unterschreitet und somit eine selbstkonsistente Lösung gefunden ist.

Zur vollständigen Spezifizierung des Kohn-Sham-Verfahrens fehlt noch ein Ausdruck für E_{xc} (siehe z.B. Ref. [142] für einen Überblick). Kohn und Sham approxmierten E_{xc} über die Annahme eines lokal homogenen Elektronengases [143]. Eine genauere Beschreibung der Austauschwechselwirkung erhält man durch Berücksichtigung von Gradienten des inhomogenen Elektronengases [139]. Durch explizites Berechnen der Hartree-Fock-Austauschwechselwirkung [144] kann die Genauigkeit weiter erhöht werden, allerdings zum Preis eines noch größeren Rechenaufwands. In dieser Arbeit werden die Kombinationen BP, BLYP und B3LYP aus dem gradienten-korrigierten Austauschfunktional von Becke (B) [145], bzw. dessen um Hartree-Fock-Austausch erweiterte Weiterentwicklung B3 [146, 147] und den Korrelationsfunktionalen von Perdew (P) [148] bzw. Lee, Yang und Parr (LYP) [149] verwendet, und u.a. deren Effekt auf die Qualität molekularer Schwingungsspektren untersucht.

Konkrete Implementierungen des Kohn-Sham-Verfahrens stellen die Einelektron-Orbitale durch Basissätze dar, deren endliche Größe ebenfalls einen Kompromiss aus Genauigkeit und Effizienz darstellt [77]. Neben an den Kernorten lokalisierten Gauß'schen Basisfunktionen [150, 151] werden auch ebene Wellen (z.B. im gitterbasierten DFT-Programm CPMD [152]) oder Mischformen [153–155] als Basis verwendet. Um die Zahl der explizit zu behandelnden Elektronen zu reduzieren, ist es in auf ebenen Wellen basierenden Verfahren üblich, kernnahe Elektronen durch Pseudopotentiale zu modellieren [77]. In dieser Arbeit werden hauptsächlich die normerhaltenden Pseudopotentiale von Troullier und Martins (MT) verwendet [156].

Die im Falle der Born-Oppenheimer-MD (vgl. Abschnitt 1.2) bei Verwendung eines DFT-Modells in jedem Integrationsschritt nötige aufwändige Selbstkonsistenziteration der Elektronendichte wird durch eine von Car und Parrinello vorgeschlagene Alternative, der CP-MD, vermieden [77, 157]. Durch Einführung einer fiktiven Elektronenmasse und einer damit ermöglichten adiabatischen Elektronendynamik können die Elektronen bei Bewegung der Kerne nahe am Grundzustand gehalten werden.

Eignung der DFT zur Berechnung von Schwingungsspektren

Die Bedeutung der DFT zeigt sich unter anderem in der Verleihung des Chemie-Nobelpreises 1998 für Walter Kohn für die Entwicklung der DFT, und an John Pople für seine Beiträge zur Entwicklung und Verbreitung von Quantenchemie-Software [158].

Für die Behandlung IR-spektroskopischer Probleme bietet die DFT schon bei der Verwendung von Funktionalen mit moderatem Rechenaufwand (wie z.B. BP) die für einen

Vergleich mit experimentellen Daten erforderliche Genauigkeit, wie durch Normalmodenanalysen am Beispiel von *p*-Benzochinon [159], und einer Reihe anderer organischer und anorganischer kleiner, isolierter Moleküle [160, 161] bzw. kleinen Molekülkomplexen [24] gezeigt wurde. Aufgrund der in Abschnitt 1.1.2 beschriebenen solvatochromen Frequenzverschiebungen kann das für isolierte Moleküle berechnete Schwingungsspektrum aber nur schwer mit in Lösung gemessenen Spektren verglichen werden [27].

Trotz der durch CP-MD [157] ermöglichten Effizienzsteigerung ist eine Beschreibung der Dynamik von Molekül-Lösungsmittelkomplexen mit DFT aufgrund des weiterhin hohen Rechenaufwands auch mit heutigen Supercomputern auf kleine Längen- und Zeitskalen begrenzt. Entsprechend beschränken sich solche Anwendungen auf kurze Referenzsimulationen von einigen Dutzend Pikosekunden Dauer, und auf kleine Modellmoleküle wie NMA [59, 162], Alanindipeptid [163] oder Phosphationen [164] in nur wenig ausgedehnten periodischen Lösungsmittelumgebungen oder auf pures Wasser [165–169].

1.2.3 QM/(P)MM-Hybridverfahren

Damit kann das in dieser Arbeit behandelte grundlegende Problem der Berechnung von Schwingungsspektren aus MD-Simulationen folgendermaßen formuliert werden: Die im Abschnitt 1.2.2 vorgestellte DFT liefert die erforderliche Genauigkeit, der hohe Rechenaufwand macht aber die Behandlung größerer Moleküle in einer ausgedehnten Lösungsmittelumgebung unmöglich. Einfache (P)MM-Kraftfelder, wie sie im Abschnitt 1.2.1 skizziert sind, erlauben zwar die effiziente Behandlung von Systemen solcher Größe, können aber die spektroskopischen Eigenschaften nur unzureichend beschreiben.

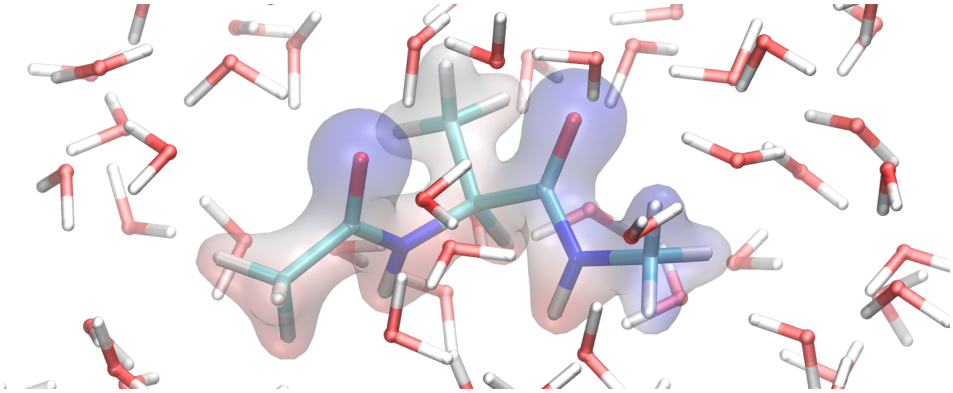


Abbildung 1.7: Beispiel für ein QM/MM-Hybridmodell: Alanindipeptid (DFT) in Wasser (MM). Die Elektronenverteilung des durch DFT beschriebenen Moleküls wird explizit berechnet und ist hier durch eine Isofläche angedeutet.

Einen Weg zur Lösung dieses Problems lieferten die anfangs erwähnten Hybridmodelle von Warshel und Levitt. In deren wegweisenden Arbeit [31] untersuchten sie eine chemische Reaktion im Enzym Lysozym, die eine quantenmechanische Beschreibung des reaktiven Zentrums erforderte. Erstmalig wurde aber dessen chemisch träge Protein- und Wasserumgebung nicht einfach ignoriert, sondern mittels einer PMM-Beschreibung, in der sowohl die Proteinatome als auch die Wasserumgebung explizit polarisierbar waren,

berücksichtigt. Mit Hilfe dieses hybriden QM/PMM-Modells konnten alle für die untersuchte enzymatische Reaktion wichtigen physikalischen Effekte, wenn auch teilweise grob genähert, berücksichtigt werden.

In dieser Studie [31] wurden viele wesentliche Konzepte entwickelt, die auch heute noch bei der Konstruktion und Anwendung solcher QM/(P)MM-Hybridmodelle, für die Abbildung 1.7 ein typisches Beispiel zeigt, von Bedeutung sind. So verwendeten die Autoren erstmalig ein polarisierbares Kraftfeld [101] zur Modellierung der Proteinumgebung durch induzierbare Punktdipole an den Atomorten. Die weiter vom reaktiven Zentrum entfernte Wasserumgebung wurde mit einem vereinfachten Polarisationsmodell vergröbert behandelt. Dieser dreistufige Multiskalenansatz (QM-Region, Protein, Wasser) erweiterte die erreichbare Systemgröße entscheidend, sodass auch langreichweitige dielektrische Effekte auf das reaktive Zentrum erfasst werden konnten. Des Weiteren wurde eine erste einfache Technik für das Schneiden von kovalenten Bindungen zwischen dem QM- und dem PMM-Fragment verwendet, und die van der Waals-Wechselwirkungen zwischen QM- und PMM-Atomen durch empirische Lennard-Jones-Potentiale modelliert (vgl. Abschnitt 1.2.1). Die verwendete Form der QM/PMM-Energiefunktion [31]

$$U^{\mathrm{QM/PMM}} = U^{\mathrm{QM}} + U^{\mathrm{PMM}} + U_{\mathrm{elec}}^{\mathrm{QM/PMM}} + U_{\mathrm{vdW}}^{\mathrm{QM/PMM}} + U_{\mathrm{b}}^{\mathrm{QM/PMM}}, \tag{1.16}$$

in der U^{QM} und U^{PMM} [vgl. Glg. (1.6)] die potentielle Energie des QM- bzw. PMM-Fragments bezeichnet, und die zusätzlich die elektrostatischen ($U_{\mathrm{elec}}^{\mathrm{QM/PMM}}$), van der Waals ($U_{\mathrm{vdW}}^{\mathrm{QM/PMM}}$) und die gebundenen Wechselwirkungen ($U_{\mathrm{b}}^{\mathrm{QM/PMM}}$) zwischen QM- und PMM-Fragment beinhaltet, liegt im wesentlichen auch den modernen QM/(P)MM-Methoden zugrunde.

Wie in zahlreichen Übersichtsartikeln dargestellt ist [170–174], wurde Gleichung (1.16) seitdem auf diverse Arten implementiert. Die resultierenden QM/(P)MM-Verfahren unterscheiden sich u.a. in der Wahl der QM-Methode, der Beschreibung der nicht-gebundenen QM/(P)MM-Wechselwirkungen, oder der Behandlung kovalenter Bindungen zwischen den QM und (P)MM-Fragmenten. Das von Warshel und Levitt ursprünglich verwendete PMM-Kraftfeld wurde allerdings häufig durch ein einfacheres, unpolarisierbares MM-Kraftfeld ersetzt. Allen diesen Implementierungen liegt die Suche nach einem auf eine bestimmte Fragestellung hin optimierten Kompromiss aus Genauigkeit und Recheneffizienz zugrunde. Da der Untersuchungsgegenstand der vorliegenden Arbeit die theoretische Beschreibung von IR-Schwingungsspektren ist, beschränken wir uns im folgenden auf ein DFT-basiertes Hybridverfahren (vgl. Abschnitt 1.2.2).

Das DFT/MM-Verfahren von Eichinger et al.

Eichinger *et al.* stellten 1999 eine DFT/MM-Hybridmethode vor, die speziell für die DFT-Berechnung der IR-Schwingungsspektren von Molekülen in ausgedehnten MM-Lösungsmittelumgebungen konzipiert wurde [33]. Die Implementierung verwendete das gitterbasierte DFT-Programm CPMD [152], das mit dem MM-MD Program EGO [175] über das Dateisystem kommunizierte [176]. Die in jedem Integrationsschritt für die Auswertung von Gleichung (1.12) nötige Berechnung des durch die Partialladungen q_i des

MM-Fragments auf dem ausgedehnten DFT-Gitter ($\gtrsim 10^6$ Gitterpunkte) erzeugten äußeren Potentials $\Phi(\mathbf{r} \,|\, q_i, \mathbf{r}_i)$ [Glg. (1.5)] wurde durch eine Erweiterung des SAMM-Algorithmus [123–125] effizient ermöglicht — wie bei Warshel und Levitt [31] wurde also ein echter Multiskalenansatz verwendet. Um artifizielle Störungen der DFT-Elektronendichte durch die bei kleinen Abständen divergierenden Coulombpotentiale (1.5) zu vermeiden, mussten die Punktladungen des umgebenden MM-Fragments gaußförmig verschmiert werden. Für die Berechnung der vom DFT-Fragment auf die MM-Atome ausgeübten Kräfte wurde eine Partialladungsnäherung für die DFT-Atome verwendet, und die resultierende Verletzung des Newton'schen Reaktionsprinzips durch eine Kraftkorrektur ausgeglichen. Ein SPLAM (für *scaled position link atom method*) genanntes Verfahren erlaubte die akkurate Behandlung von kovalenten Bindungen zwischen DFT- und MM-Fragmenten durch Einführung eines zusätzlichen sog. Link-Atoms. Testrechnungen zeigten, dass die neue Methode erstmals offenbar stabile DFT/MM-MD-Simulationen in relativ großen MM-Umgebungen sowie die Berechnung der IR-Spektren von kovalent an das MM-Fragment gebundenen DFT-Fragmenten ermöglichte [33].

Es folgte eine Vielzahl von Anwendungen zu unterschiedlichen Fragestellungen [21, 25, 34–37, 50, 63, 137, 177–183]. So wurden mit DFT/MM die IR-Spektren von im photosynthetischen Reaktionszentrum des Purpurbakteriums *Rhodobacter Sphaeroides* eingebetten Chinon-Farbstoffen berechnet und Fehler in der aus Kristallographie abgeleiteten Proteinstruktur aufgedeckt [34]. Ferner wurden die Schwingungsspektren von *p*-Benzochinon [177], eines lichtschaltbaren β-Hairpin-Moleküls [21], von AGn [63] und von Phosphationen [35] — jeweils in wässriger Lösung — untersucht. Studien der IR-Spektren von Triphosphatgruppen im Enzym *RAS p21* [178, 179] sowie des Retinalchromophors [36, 137] in der Bindungstasche der lichtgetriebenen Protonenpumpe *Bacteriorhodopsin* umgaben das jeweilige DFT-Fragment mit komplexeren MM-Umgebungen. Weiterhin wurden die IR-Spektren von Flavin-Farbstoffen in unterschiedlichen chemischen Zuständen in Wasser [25, 182] und in deren nativer Umgebung, den sog. BLUF-Domänen [37], berechnet und ermöglichten so die Zuordnung experimentell gemessener IR-Absorptionsbanden zu molekularen Schwingungsmoden sowie die Identifizierung fehlerhafter Protein-Kristallstrukturen. Auch grundlegende elektrostatische Eigenschaften von Wassermolekülen in flüssiger Phase waren Ziel von DFT/MM-Studien [50, 183].

In den eben zitierten DFT/MM-Studien wurden Schwingungsspektren fast ausnahmslos über die *instantane Normalmodenanalyse* (INMA) des DFT-Fragments in eingefrorenen MM-Solvatstrukturen berechnet [177, 180, 181]. Ein alternatives Verfahren beruht auf der Fouriertransformation der Zeit-Autokorrelation des molekularen Dipolmoments (FTTCF, für *Fourier transformation of the time correlation function*) [181], berücksichtigt Quantenkorrekturen [184] zu den klassisch behandelten Kernbewegungen, und gewinnt das Spektrum direkt aus einer ausgedehnten MD-Trajektorie. Im Gegensatz zu INMA kann FTTCF daher dynamische Effekte erfassen, und so z.B. die Form von Absorptionsmaxima wesentlich besser beschreiben [27, 181]. Zusammen mit der Analyse der Trajektorie über *generalisierte Normalkoordinaten* (engl. *generalized normal coordinates*, GNC) [185, 186] kann, bei ausreichend großer Datenbasis, das Schwingungsspektrum in die zugrundeliegenden Normalmoden (vgl. Abschnitt 1.1.2) zerlegt werden. FTTCF und GNC sind daher die Methoden, mit denen Spektren von DFT/MM-Modellen vorzugs-

weise berechnet werden sollten. Die dafür erforderlichen ausgedehnten DFT/MM-MD Trajektorien stellen aber hohe Anforderungen an die Stabilität und die Recheneffizienz der DFT/MM-Methode.

Notwendige Erweiterungen des DFT/MM-Verfahrens

Die bereits erwähnte Verletzung des Reaktionsprinzips durch die von Eichinger *et al.* verwendete Partialladungsnäherung für die DFT-Atome [33] verursachte in einigen MD-Simulationen größere Dynamikartefakte als ursprünglich erwartet [21, 187]. Alternative DFT/MM-Kopplungsschemata zeigten, dass diese durch eine Hamilton'sche Formulierung der DFT/MM-Wechselwirkungen, in der alle atomaren Kräfte als Ableitungen einer einzigen DFT/MM-Energiefunktion $U^{\mathrm{DFT/MM}}$ [vgl. Glg. (1.16)] berechnet werden, vermieden werden können [188, 189].

Ein weiteres Manko von Eichingers Methode deckten die Studien der in Wasser gelösten Phosphationen [35] sowie der in Proteinumgebungen eingebetteten Chromophore [36, 37] auf: die DFT/MM-Methode vernachlässigte, wie viele andere QM/MM-Verfahren [170, 172] (und im Gegensatz zum ursprünglichen Vorschlag von Warshel und Levitt [31]) die expliziten elektronischen Polarisationseffekte im MM-Fragment des Hybridsystems. Im Falle der in Proteine eingebetteten Farbstoffe mussten diese Effekte durch eine aufwändige, aus DFT/MM-Rechnungen abgeleitete und spezifisch auf die Umgebung des jeweiligen Chromophors angepasste iterative Reparametrisierung der Partialladungen des MM-Kraftfelds berücksichtigt werden [36, 37]. Für einfach und doppelt geladene Phosphationen in Wasser wurde eine Unterschätzung der solvatochromen Verschiebungen festgestellt [35], für die als Ursachen eine fehlerhafte Nahordnung der MM-Wassermoleküle um das DFT-Fragment [164], und darüber hinaus die fehlende Polarisierbarkeit des verwendeten MM-Wassermodells [66] in Frage kommen.

Außerdem stellten sich die bei Verwendung von punktförmigen MM-Partialladungen für die elektrostatische Wechselwirkung mit dem DFT Fragment notwendigen atomaren Glättungsskalen als für die Nahstruktur der Verteilung der MM-Moleküle um das DFT-Fragment kritische Parameter heraus [188]. Dieser Befund stellte die Verwendung von MM-Punktladungen in DFT/MM-Modellen grundsätzlich in Frage.

Die Erweiterung der DFT/MM-Hybridmethode von Eichinger *et al.* [33] zu einem Hamilton'schen DFT/PMM-Verfahren, das die vom DFT-Fragment auf die PMM-Atome ausgeübten Kräfte als Hellmann-Feynman-Kräfte [190] berechnet und so stabile MD Simulationen von in PMM-Umgebungen eingebetteten DFT-Fragmenten ermöglichen sollte, war daher der logische nächste Schritt in der Methodenentwicklung. Erste Grundlagen wurden hier in der Diplomarbeit von Benedikt Breitenfeld gelegt [187]. Die resultierende frühe Version des induzierbare Dipole einsetzenden DFT/PMM-Verfahrens war aber auf die Simulation kleiner Systeme wie hybride Wasserdimere beschränkt, da die effiziente Hamilton'sche Einbettung [188] des DFT-Fragments in den SAMM-Algorithmus noch ausstand. Im Rahmen meiner Masterarbeit [191] wurde dann die dafür nötige Erweiterung des zur damaligen Zeit aktuellen SAMM$_p$-Algorithmus [126] entwickelt.

1.3 Zielsetzung und Aufbau der Arbeit

Die Entwicklung einer hocheffizienten Hamilton'schen DFT/PMM-Methode, sowie die Beantwortung wichtiger methodischer Fragen zur physikalisch korrekten und genauen Beschreibung der MD von (Poly-)Peptiden speziell für die Berechnung deren IR-Schwingungsspektren war das Ziel dieser Dissertation, die ich in der Arbeitsgruppe für theoretische molekulare Biophysik am Lehrstuhl für BioMolekulare Optik der Ludwigs-Maximilians-Universität München angefertigt habe. Die Finanzierung des Projekts erfolgte einerseits im Rahmen des Sonderforschungsbereich 749/C4 der Deutschen Forschungsgemeinschaft, der die Weiterentwicklung und prototypische Erprobung von Simulationsmethoden zur Beschreibung der Dynamik (bio)chemischer Grundzustandsprozesse in kondensierter Phase zum Ziel hatte, sowie andererseits durch das *Kompetenznetzwerk für Wissenschaftliches Höchstleistungsrechnen in Bayern* (KONWIHR-III) des Bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst. Alle im Rahmen dieser Arbeit entwickelten Simulationstechniken sind im PMM-MD Programm IPHIGENIE [192], dem Nachfolger von EGO, implementiert, das unter der Leitung von Gerald Mathias der wissenschaftlichen Öffentlichkeit zugänglich gemacht worden ist. Die vier dieser kumulativen Dissertation zugrundeliegenden Artikel sind im Kapitel 2 in den Abschnitten 2.1 bis 2.4 nachgedruckt.

Aufbauend auf den oben beschrieben Vorarbeiten [187, 191] konnte die Hamilton'sche DFT/PMM-Hybridmethode erfolgreich zur Publikationsreife gebracht werden. Die entsprechende Veröffentlichung [112], die in Abschnitt 2.1 abgedruckt ist, stellt im Detail die Integration des DFT-Fragments in den FMM-Algorithmus SAMM$_p$ [126] vor. Zur Ermöglichung Hamilton'scher Dynamik war einerseits die symmetrische Behandlung der Auswertung des durch das PMM-Fragment generierten Potentials auf dem DFT-Gitter und der Berechnung der von der DFT-Ladungsdichte auf die PMM-Atome ausgeübten Hellmann-Feynman-Kräfte [190] nötig. Ferner musste der Algorithmus zur Nachführung des DFT-Gitters an Translationen des DFT-Fragments überarbeitet werden. Um den zusätzlichen Rechenaufwand gegenüber DFT/MM-Rechnungen gering zu halten, wurde ein mehrstufiges DFT-Konvergenzkriterium sowie die Extrapolation der PMM-Dipole eingeführt.

Erste Testanwendungen an größeren DFT-Fragmenten wie Alanindipeptid zeigten allerdings, dass die Genauigkeit des auf den fixen SAMM$_p$-Distanzklassen [126] basierenden DFT/PMM-Verfahrens [112] kritisch und molekülspezifisch von der Wahl dieser Klassen abhing. Entsprechend wurden im Rahmen der in Abschnitt 2.2 abgedruckten Veröffentlichung [193] die mittlerweile zur Verfügung stehende SAMM$^\chi$-Methode [109], die die tatsächlichen Größe der den FMM-Entwicklungen zugrundeliegenden Ladungsclustern berücksichtigt [vgl. Glg. (1.10)] und so massive Genauigkeits- und Effizienzgewinne versprach, auf die DFT/PMM-Kopplung angewendet.

Ferner limitierte die Tatsache, dass das MPI-parallelisierte PMM-MD Programms IPHIGENIE [192] in DFT/PMM-Hybridrechnungen noch auf einen Rechenkern beschränkt war, die Anwendbarkeit der neuen DFT/PMM-Methode erheblich. Ein weiteres im Rahmen von Ref. [193] verfolgtes Ziel war daher die vollständige MPI/OpenMP-Paralle-

lisierung der DFT/PMM-Implementierung sowie die Integration von CPMD [152] in IPHIGENIE [192], um den Einsatz von IPHIGENIE/CPMD auf Großrechnern zu ermöglichen.

Bevor das dann mit dem Abschluß der technischen Entwicklungen verfügbare, hoch leistungsfähige Programmpaket IPHIGENIE/CPMD allerdings zur Berechnung der Schwingungsspektren von (Poly-)Peptiden eingesetzt werden konnte, waren noch einige weitere Vorarbeiten nötig. So waren zwar mithilfe der DFT/PMM-Technologie inzwischen polarisierbare Wassermodelle entwickelt worden [55, 129], da diese jedoch die elektrostatische Signatur eines Wassermoleküls durch Punktladungen modellierten, waren sie für die Anwendung als Lösungsmittel in DFT/PMM-MD Simulationen ungeeignet.

Diese offenen Fragen wurden in der Veröffentlichung [58], die in Abschnitt 2.3 zu finden ist, beantwortet. Das dort nach der in Ref. [55] und Ref. [129] entwickelten Strategie konstruierte *Gauß'sche polarisierbare Sechspunktmodell* (GP6P) für Wasser sollte sich durch die ausschließliche Verwendung gaußförmig verschmierter Quellen des elektrostatischen Potentials speziell als Lösungsmittel in DFT/PMM-Anwendungen eignen. Es wurde untersucht, mit welcher Qualität dieses Wassermodell wichtige Eigenschaften flüssigen Wassers beschreiben kann. Anschließend mussten die bisher unbekannten Parameter der Lennard-Jones-Wechselwirkung zwischen AGn und GP6P-Molekülen optimiert werden. Da diese Parameter die Nahordnung der PMM-Wassermoleküle um die AGn, und damit deren Polarisation steuern, hat ihre Wahl auch Einfluss auf die Qualität der Beschreibung solvatochromer Effekte in den Schwingungsspektren der AGn (vgl. Abschnitt 1.1.2). Die Parameteroptimierung erfolgte über eine Anpassung der radialen Nahstruktur des PMM-Wassers um ein durch DFT beschriebenes NMA-Molekül an eigens mittels DFT-MD generierte Referenzdaten. Zur Verringerung des Rechenaufwands wurde für diesen Zweck eine effiziente DFT/PMM-mean-field-Methode entwickelt. Schließlich wurden die solvatochromen Verschiebungen durch wässrige Lösung (vgl. Abb. 1.4) sowie Isotopeneffekte durch DFT-MD und DFT/PMM-MD Rechnungen am isolierten bzw. in GP6P gelösten NMA untersucht und mit experimentellen Daten verglichen [10].

Um ein effizientes Abtasten von Peptidkonformationen auch für DFT/PMM-Modelle zu ermöglichen, untersucht die in Abschnitt 2.4 abgedruckte Veröffentlichung [194] schließlich, wie die von Denschlag *et al.* [195] entwickelte SST-Methode (für *simulated solute tempering*) auch auf polarisierbare Simulationssysteme wie PMM oder DFT/PMM angewendet werden kann. Es wird eine einfache Vorschrift entwickelt, mit der die für DFT/PMM-SST-MD Simulationen notwendigen SST-Gewichtsparameter aus deutlich weniger aufwändigen PMM-SST-Rechnungen abgeleitet werden können. Das Verfahren wird dann erfolgreich zur Berechnung der Konformationslandschaft eines DFT-Modells von Alanindipeptid in wässriger PMM-Lösung angewendet.

Kapitel 3 fasst die Ergebnisse der Arbeit kurz zusammen und liefert einen Ausblick auf zukünftige algorithmische Weiterentwicklungen und Anwendungen.

2 Entwicklung einer hocheffizienten Hamilton'schen DFT/PMM-Methode

Die vier in diesem Kapitel abgedruckten und im *Journal of Chemical Physics* bzw. im *Journal of Chemical Theory and Computation* veröffentlichten Artikel stellen die grundlegende Hamilton'sche Kopplung des DFT-Fragments mit der PMM-Umgebung dar (Abschnitt 2.1), und erklären die effiziente Einbettung in die SAMM-Hierarchie (Abschnitt 2.2). Ferner dokumentieren sie die Entwicklung des GP6P-Wassermodells und die Berechnung der IR-Spektren von NMA in wässriger Lösung mit DFT/PMM-MD (Abschnitt 2.3). Schließlich stellen sie die zur effizienten Abtastung von Konformationsräumen hilfreiche DFT/PMM-SST-Methode vor (Abschnitt 2.4).

2.1 Eine Hamilton'sche DFT/PMM-Kopplung

In der nachfolgend abgedruckten Publikation

Coupling DFT to polarizable force fields for efficient and accurate Hamiltonian molecular dynamics simulations

Magnus Schwörer, Benedikt Breitenfeld, Philipp Tröster, Sebastian Bauer, Konstantin Lorenzen, Paul Tavan und Gerald Mathias

J. Chem. Phys. **138**, 244103 (2013),

die ich zusammen mit Benedikt Breitenfeld, Philipp Tröster, Sebastian Bauer, Konstantin Lorenzen, Paul Tavan und Gerald Mathias verfasst habe, wird die Erweiterung von IPHIGENIE um induzierbare Gauß'sche Dipole, die Hamilton'sche Einbettung des DFT-Fragments in den damals zur Verfügung stehenden SAMM$_p$-Algorithmus [126], und eine effiziente Strategie zum Erreichen der gleichzeitigen Selbstkonsistenz von PMM- und DFT-Fragment beschrieben.

THE JOURNAL OF CHEMICAL PHYSICS **138**, 244103 (2013)

Coupling density functional theory to polarizable force fields for efficient and accurate Hamiltonian molecular dynamics simulations

Magnus Schwörer, Benedikt Breitenfeld, Philipp Tröster, Sebastian Bauer, Konstantin Lorenzen, Paul Tavan, and Gerald Mathias[a)]
Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

(Received 5 April 2013; accepted 3 June 2013; published online 25 June 2013)

Hybrid molecular dynamics (MD) simulations, in which the forces acting on the atoms are calculated by grid-based density functional theory (DFT) for a solute molecule and by a polarizable molecular mechanics (PMM) force field for a large solvent environment composed of several 10^3–10^5 molecules, pose a challenge. A corresponding computational approach should guarantee energy conservation, exclude artificial distortions of the electron density at the interface between the DFT and PMM fragments, and should treat the long-range electrostatic interactions within the hybrid simulation system in a linearly scaling fashion. Here we describe a corresponding Hamiltonian DFT/(P)MM implementation, which accounts for inducible atomic dipoles of a PMM environment in a joint DFT/PMM self-consistency iteration. The long-range parts of the electrostatics are treated by hierarchically nested fast multipole expansions up to a maximum distance dictated by the minimum image convention of toroidal boundary conditions and, beyond that distance, by a reaction field approach such that the computation scales linearly with the number of PMM atoms. Short-range over-polarization artifacts are excluded by using Gaussian inducible dipoles throughout the system and Gaussian partial charges in the PMM region close to the DFT fragment. The Hamiltonian character, the stability, and efficiency of the implementation are investigated by hybrid DFT/PMM-MD simulations treating one molecule of the water dimer and of bulk water by DFT and the respective remainder by PMM. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4811292]

I. INTRODUCTION

In a seminal paper devoted to the study of enzymatic reactions, Warshel and Levitt[1] introduced in 1976 a quantum-classical coupling scheme for a molecule, which is described by quantum mechanics (QM) and is embedded in a condensed phase environment modeled by a polarizable molecular mechanics (PMM) force field. In their abstract, these authors emphasized that the "solvation energy resulting from this polarization is considerable and must be included in any realistic calculation" of molecules in condensed phase.

However, as documented in a recent review on QM/MM methods for biomolecular systems,[2] this advice was subsequently ignored in most applications. Instead so-called standard MM force fields like AMBER,[3] CHARMM,[4] OPLS-AA,[5] or GROMOS[6] were generally applied to the MM part of hybrid simulation systems. These force fields model the electrostatic signatures of molecules or of molecular fragments by static partial charges localized at the atoms and, therefore, can account for the effects of electronic polarization only by the mean field approximation, which is highly questionable for inhomogeneous and non-isotropic biomolecular systems.[7] There are notable exceptions which combined a polarizable force field for the MM fragment with semi-empirical quantum chemistry for the QM fragment.[8–12] Combinations of higher-level QM treatments (density functional theory[13,14] (DFT) or *ab initio* quantum chemistry) with PMM force fields were either restricted to the energetics of static systems,[15–23] to small molecular clusters,[24–29] or describe the dynamics only in parts of the simulation system.[30,31] Other approaches augment DFT atoms with self-consistent polarization terms (SCP-DFT) to correct the deficiencies of the long-range electrostatics and dispersion description within certain exchange-correlation functionals.[32,33]

The development of hybrid methods combining grid-based DFT with non-polarizable MM force fields started with the work of Eichinger *et al.*,[34] which particularly aimed at accurate computations of vibrational spectra of molecules in condensed phase environments from hybrid MD simulations. Since then corresponding applications have demonstrated the power of this approach.[35–37] Subsequently, two further DFT/MM implementations[38,39] took up the challenge posed by the requirement to combine DFT treatments of a molecule in an efficient and accurate way with large scale MM environments. Here, Laio *et al.*[38] emphasized the need of a fully Hamiltonian description, which was violated by certain approximations applied by Eichinger *et al.*,[34] while Laino *et al.*[39] additionally provided a clever suggestion for the efficient computation of the electrostatic interaction between the DFT and MM fragments.

However, applications of the above DFT/MM setting to the computation of infrared (IR) spectra of biological chromophores like retinal in bacteriorhodopsin[40,41] or flavin in blue light sensing domains[42] through instantaneous normal

a)Electronic mail: gerald.mathias@physik.uni-muenchen.de

mode analyses[36,43] also revealed those limitations, which are due to the neglected polarizability of the MM protein environments surrounding the DFT chromophores. The corresponding errors in the computed vibrational spectra could be largely removed by iterative DFT/MM calculations of polarized force fields in the respective chromophore binding pockets, thus, uniquely proving that the neglected polarizabilities were the main cause of the earlier ill-descriptions.

A similar attempt to compute the vibrational spectra of phosphate ions in aqueous solution[44] showed that the DFT/MM calculations largely underestimate the solvatochromic shifts in the IR spectra. Here, these underestimates were erroneously attributed to the neglected polarizability of the solvating water, which had been modeled by Jorgensen's "transferable three point interaction potential" (TIP3P).[45] By contrast, recent "first principles" DFT-MD simulations of phosphates in small periodic water boxes have clearly shown[46] that the use of the TIP3P model entailed highly erroneous structures for the first solvation shell, which are mainly due to its simplified structure and to a lesser degree due to the neglected polarizability. Hence, it remains to be seen whether improved (and polarizable) MM water models combined with a DFT description of the phosphate solutes can predict the solvatochromic shifts in the phosphate IR spectra at a quality comparable to that of the very expensive "first principles" DFT-MD simulations.

To enable rapid and nevertheless accurate computations of solvatochromic effects in chromophore IR spectra, the construction of a new and efficient DFT/PMM implementation therefore seemed necessary.

In this paper, we address two issues. First, we thoroughly revise the DFT/MM suggestion made by Eichinger *et al.*[34] and develop an efficient, accurate, and fully Hamiltonian electrostatic DFT/MM coupling scheme whose computational effort scales logarithmically with the number of condensed phase atoms surrounding the DFT fragment. Preserving the thereby achieved levels of accuracy and efficiency, we next extend this scheme by including dynamic polarization effects through inducible atomic dipoles. We give analytical expressions for the calculation of the forces and, therefore, are able to employ the new DFT/PMM scheme for molecular dynamics simulations. For the implementation, the program packages of choice are the parallelized PMM-MD program IPHIGENIE[47] and the parallelized grid-based plane wave DFT program CPMD.[48]

II. THEORY

The Hamiltonian of a DFT/PMM hybrid system can be decomposed into the following four contributions:

$$H = H_{\mathrm{MM}} + H_{\mathrm{PMM}} + H_{\mathrm{DFT}} + H_{\mathrm{DFT/(P)MM}}. \tag{1}$$

Here, H_{MM} represents one of the standard MM force fields[3–6] including the kinetic energy of the atoms, and

$$H_{\mathrm{PMM}} = \frac{1}{2}\sum_{i,j\neq i} q_i \Phi(\mathbf{r}_i \,|\, \mathbf{p}_j, \mathbf{r}_j, \tilde{\sigma}_j) - \frac{1}{2}\sum_i \mathbf{p}_i \cdot \langle \mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i)\rangle_{\tilde{\sigma}_i} + \frac{1}{2}\sum_i \mathbf{p}_i^2/\alpha_i \tag{2}$$

accounts[1,49–51] for the energy contribution of polarizable Gaussian dipoles

$$\mathbf{p}_i^{\mathrm{G}}(\mathbf{r} \,|\, \mathbf{r}_i, \tilde{\sigma}_i) = \mathbf{p}_i \, g(\mathbf{r} \,|\, \mathbf{r}_i, \tilde{\sigma}_i)$$

of strengths $\mathbf{p}_i$ and widths $\tilde{\sigma}_i$, which are located at the atomic positions $\mathbf{r}_i$ and have the shape functions

$$g(\mathbf{r} \,|\, \mathbf{r}_i, \tilde{\sigma}_i) = \frac{1}{\left(2\pi\tilde{\sigma}_i^2\right)^{3/2}} \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{2\tilde{\sigma}_i^2}\right]. \tag{3}$$

Note that in PMM force fields, the use of Gaussian dipoles yields an enhanced algorithmic stability,[52,53] if the widths $\tilde{\sigma}_i$ are chosen sufficiently large, i.e., typically $\tilde{\sigma}_i \approx 0.1$ nm.

The symbol $\Phi(\mathbf{r}_i \,|\, \mathbf{p}_j, \mathbf{r}_j, \tilde{\sigma}_j)$ in Eq. (2) denotes the electrostatic potential generated at the position $\mathbf{r}_i$ of an atom $i \neq j$ by a Gaussian dipole $\mathbf{p}_j^{\mathrm{G}}(\mathbf{r} \,|\, \mathbf{r}_j, \tilde{\sigma}_j)$. Furthermore, the bracket expression

$$\langle f(\mathbf{r}_i)\rangle_{\tilde{\sigma}_i} \equiv \int f(\mathbf{r}) g(\mathbf{r} \,|\, \mathbf{r}_i, \tilde{\sigma}_i) d\mathbf{r} \tag{4}$$

denotes the average of a function $f(\mathbf{r})$ over the volume occupied by $g(\mathbf{r} \,|\, \mathbf{r}_i, \tilde{\sigma}_i)$. If $\mathbf{E}(\mathbf{r} \,|\, q_j, \mathbf{r}_j) = -\nabla(q_j/|\mathbf{r} - \mathbf{r}_j|)$ is the field of a point charge q_j and $\mathbf{E}(\mathbf{r} \,|\, \mathbf{p}_j, \mathbf{r}_j, \tilde{\sigma}_j) = -\nabla\Phi(\mathbf{r} \,|\, \mathbf{p}_j, \mathbf{r}_j, \tilde{\sigma}_j)$ is the field of a Gaussian dipole at $\mathbf{r}_j$, then

$$\langle \mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i)\rangle_{\tilde{\sigma}_i} \equiv \sum_{j\neq i} \left\langle \mathbf{E}(\mathbf{r}_i \,|\, q_j, \mathbf{r}_j) + \mathbf{E}(\mathbf{r}_i \,|\, \mathbf{p}_j, \mathbf{r}_j, \tilde{\sigma}_j)\right\rangle_{\tilde{\sigma}_i} \tag{5}$$

is the field polarizing atom i. Assuming linear response, the dipole strengths $\mathbf{p}_i$ are calculated by

$$\mathbf{p}_i = \alpha_i \langle \mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i)\rangle_{\tilde{\sigma}_i} \tag{6}$$

from the scalar atomic polarizabilities α_i and from the polarizing fields (5) in a self-consistent field iteration (PMM-SCF).[54,55] The last term in Eq. (2) is the self-energy required to create the dipoles $\mathbf{p}_i$. If Eq. (6) is self-consistently fulfilled, this self-energy cancels the second term in Eq. (2) and the first term remains as the polarization contribution to the total energy.

In Eq. (1), H_{DFT} is the energy function of the isolated quantum system. The DFT/(P)MM interaction energy

$$H_{\mathrm{DFT/(P)MM}} = H_{\mathrm{DFT/MM}}^{\mathrm{vdW}} + H_{\mathrm{DFT/(P)MM}}^{\mathrm{bonded}} + H_{\mathrm{DFT/(P)MM}}^{\mathrm{elec}} \tag{7}$$

has a contribution from van der Waals interactions $H_{\mathrm{DFT/MM}}^{\mathrm{vdW}}$ between MM and DFT atoms, which is calculated with the applied MM force field. If chemical bonds between the DFT and (P)MM fragments exist, a term $H_{\mathrm{DFT/(P)MM}}^{\mathrm{bonded}}$ has to be included, for which several suggestions exist.[2,34]

We will, however, focus here on chemically non-bonded PMM and DFT fragments, for which the electrostatic interaction energy

$$H_{\mathrm{DFT/(P)MM}}^{\mathrm{elec}} = \int d\mathbf{r}\rho(\mathbf{r})\Phi_{\mathrm{ext}}(\mathbf{r}) \tag{8}$$

is given by the classical expression for the energy of the DFT fragment's charge density ρ in the external potential Φ_{ext} generated by the partial charges and induced Gaussian dipoles in the PMM fragment. The DFT charge density $\rho(\mathbf{r}) = \rho_{\mathrm{e}}(\mathbf{r}) + \rho_{\mathrm{c}}(\mathbf{r})$ comprises contributions $\rho_{\mathrm{e}}(\mathbf{r})$ of the valence

electrons and $\rho_c(\mathbf{r})$ of the nuclear cores. Correspondingly, the interaction Hamiltonian

$$H^{\text{elec}}_{\text{DFT/(P)MM}} = H_e + H_c \tag{9}$$

decomposes into energies

$$H_\kappa = \int d\mathbf{r}\, \rho_\kappa(\mathbf{r})\, \Phi_{\text{ext}}(\mathbf{r}), \qquad \kappa \in \{e, c\}, \tag{10}$$

associated to the electrons (e) and nuclear cores (c) of the DFT fragment.

In the computation of those atomic forces, which are caused by an external potential $\Phi_{\text{ext}}(\mathbf{r})$, the employed DFT program CPMD[48] treats the effective core charges as Gaussian distributions $q_\mu\, g(\mathbf{r}\,|\,\mathbf{r}_\mu, \sigma_\mu)$ centered with widths σ_μ around the positions $\mathbf{r}_\mu$ of the DFT atoms μ. Thus, with definition (4) of Gaussian averages one obtains from Eq. (10) for the core Hamiltonian:

$$H_c = \sum_\mu q_\mu \langle \Phi_{\text{ext}}(\mathbf{r}_\mu)\rangle_{\sigma_\mu}. \tag{11}$$

The Gaussian averages $\langle \Phi_{\text{ext}}(\mathbf{r}_\mu)\rangle_{\sigma_\mu}$ are readily calculated for external potentials generated by Gaussian charge or dipole distributions of widths σ_i, because one solely has to replace the σ_i by the widths $\sigma_{i\mu} = (\sigma_i^2 + \sigma_\mu^2)^{1/2}$ in the respective expressions for the potentials. Note that these expressions reduce for distances $r_{i\mu} \equiv |\mathbf{r}_{i\mu}| \equiv |\mathbf{r}_i - \mathbf{r}_\mu| \gg \sigma_{i\mu}$ to the potentials of point charges and dipoles.

As a result, the Gaussian approximation $q_\mu\, g(\mathbf{r}\,|\,\mathbf{r}_\mu, \sigma_\mu)$ of the nuclear pseudo-potentials enables a speedy evaluation of H_c. By contrast, the computation of H_e requires the evaluation of $\Phi_{\text{ext}}(\mathbf{r})$ at all N_γ points γ of the grid, on which ρ_e is represented in real space by CPMD.[48] Introducing the electronic grid charges

$$q_\gamma = (V_{\text{box}}/N_\gamma)\rho_e(\mathbf{r}_\gamma), \tag{12}$$

whose sum over all grid points γ in the DFT box volume V_{box} is the total charge of all valence electrons in the DFT fragment, one can numerically evaluate the integral from Eq. (10) for the electronic contribution to the DFT/(P)MM interaction Hamiltonian (9) by the expression

$$H_e = \sum_\gamma q_\gamma\, \Phi_{\text{ext}}(\mathbf{r}_\gamma). \tag{13}$$

This approximation is valid because our way of constructing Φ_{ext} guarantees, as will be explained in more detail below, that the external potential is smooth on the spatial scale defined by the DFT grid.

A. Computational issues

The forces required for MD simulations are obtained from the Hamiltonian (1) by taking negative gradients with respect to the atomic coordinates at every time step of the numerical integration of the classical equations of motion. In a DFT/MM setting, only the Kohn-Sham wave functions have to be determined in a SCF iteration (DFT-SCF) at every time step, whereas DFT/PMM simulations additionally require a PMM-SCF procedure, which has to be properly interfaced with the DFT-SCF calculations.

The computational effort required for the four components of the Hamiltonian (1) depends on the respective sizes of the DFT and PMM fragments, on the spatial resolution of the DFT grid, and on the choice of the SCF convergence criteria. Nevertheless, for a system composed of about 10^4 PMM solvent atoms and a rather small DFT solute molecule comprising about 10 DFT atoms, one may estimate that all four components of H will pose comparable numerical tasks.

Treating, for instance, a water molecule by DFT with a 70 Ry plane wave cutoff requires $N_\gamma \approx 10^6$ grid points. Within a brute force computational approach, the evaluation of Eq. (13), which is the most expensive contribution to $H_{\text{DFT/(P)MM}}$, would lead for the PMM fragment characterized above to about 10^{10} distance calculations. The associated effort would then definitely represent the computational bottleneck. For the reduction of this effort three different suggestions exist,[34,38,39] which all utilize multi-scale concepts. Here, we will adopt and extend the suggestion by Eichinger *et al.*[34] ensuring, however, the Hamiltonian character of the resulting dynamics (cf. Laio *et al.*[38]).

Accordingly, we will treat the DFT fragment as a component of the nested hierarchy into which a simulation system with toroidal boundary conditions[57] is decomposed, if the electrostatic interactions are calculated by the combination of the pth-order "structure adapted fast multipole method" with a reaction field approach (SAMM$_p$/RF) developed in Refs. 47 and 58–60. In pure MM- or PMM-MD simulations of large systems, this fast multipole method (FMM)[61–63] enables an efficient and accurate calculation of the electrostatics, which scales linearly with the number of atoms. Due to a balanced combination of mth order multipole moments with nth order local Taylor expansions, which is expressed by the equation $p = n + m$, the electrostatic forces calculated with SAMM$_p$ exactly obey Newton's reaction principle.[47,64] Furthermore, this choice additionally guarantees a minimal computational effort for a predefined level p of accuracy ($p = 4$ is the standard of the current implementation available in IPHIGENIE). A predecessor version called SAMM had been used in Eichinger's DFT/MM approach,[34] whose fully Hamiltonian DFT/(P)MM extension will be explained below.

B. DFT/PMM with SAMM$_p$

According to Eq. (13), the external potential Φ_{ext} is imported into the DFT Hamiltonian through its evaluation at the points γ of the DFT grid. For an efficient solution of this computational task, Eichinger *et al.*[34] applied and extended (cf. Figs. 4 and 5 in Ref. 34) the distance class scheme of SAMM.[58,59,65] The extended scheme partitions the PMM environment of each DFT atom μ into disjoint distance classes C_μ^l, $l = 0, \ldots, l_{\max}$. For periodic systems, a RF continuum starts beyond the outermost distance class ($l_{\max}$) at a distance dictated by the minimum image convention.[60] Figure 1 illustrates the three innermost distance classes C_μ^l, $l = 0, 1, 2$ for an atom μ of a DFT water molecule embedded in liquid PMM water.

Figure 1 shows one PMM water molecule in each of the three classes and indicates the distances d used for their

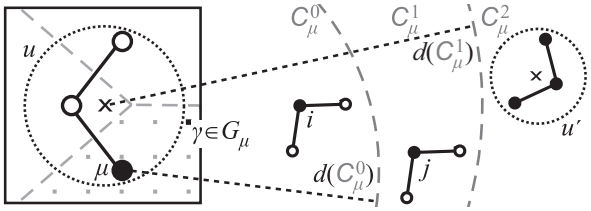


FIG. 1. Scheme of inner distance classes for the FMM evaluation of the DFT/PMM electrostatics: A water molecule (left) representing a structural unit u (dotted circle) of the SAMM$_p$ hierarchy is embedded in a rectangular DFT box, which is discretized by a grid (dots). Only those points $\gamma \in G_\mu$ are shown which belong to a selected DFT atom μ (large black dot) through the Voronoi tessellation of the box. This tessellation defines the index sets G_μ and is indicated by the dashed gray lines. Two dashed gray segments of circles [radii $d(C_\mu^0)$, $d(C_\mu^1)$] around μ and the reference point "×" of the structural unit u indicate the outer limits of the distance classes C_μ^0 and C_μ^1, respectively. Representative atoms $i \in C_\mu^0$ and $j \in C_\mu^1$ of PMM water molecules belonging to these two classes and of a structural unit $u' \in C_\mu^2$ are drawn as black dots.

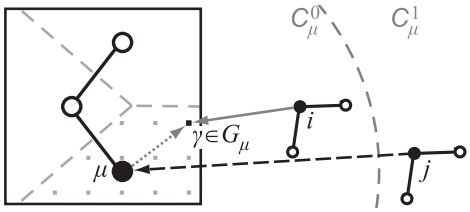


FIG. 2. Evaluation of Φ_{ext} at a grid point $\gamma \in G_\mu$: Contributions from Gaussian charges and induced dipoles of a PMM atom $i \in C_\mu^0$ are evaluated directly (solid gray arrow), whereas the contributions from more distant atoms, like the one indicated by the dashed black arrow for a PMM atom $j \in C_\mu^1$, are calculated by a Taylor expansion around the position of the DFT atom μ. The dotted gray arrow marks the connection $\mathbf{r}_{\gamma\mu}$ of the points μ and γ used in the Taylor expansion.

definition. Typical values are $d(C_\mu^0) \approx 6$ Å and $d(C_\mu^1) \approx 8$ Å. The electrostatics of the PMM atoms $i \in C_\mu^0$ is represented by Gaussian partial charges of widths σ_i and by Gaussian induced dipoles of widths $\tilde{\sigma}_i$, which all are typically smaller than 1 Å but much larger than the spacing of the DFT grid. The parameters $(\sigma_i, \tilde{\sigma}_i)$ steer the strengths of the near-field electrostatic interactions between the DFT and the PMM atoms. In the case of an aqueous PMM environment, for instance, the strength of the DFT/PMM hydrogen bonding interactions can be tuned by proper choices of these Gaussian widths.[38] Because they are about one order of magnitude smaller than the typical distances $r_{\mu j} > d(C_\mu^0)$ of atoms $j \in C_\mu^1$ from the given DFT atom μ, the Gaussian character of the PMM charges and dipoles can safely be neglected for the class C_μ^1 and beyond. In C_μ^0, the use of smoothed charge and dipole distributions is mandatory[34,38,39,56] to avoid artificial distortions of the DFT electron density $\rho_e(\mathbf{r}_\gamma)$ and to guarantee that Φ_{ext} is sufficiently smooth on the scale of the DFT grid spacing.

The Voronoi tessellation of the DFT box characterized by the gray dashed lines in Fig. 1 decomposes the whole DFT grid into disjoint subsets G_μ associated to the various DFT atoms μ. Correspondingly, the sum $\sum_\gamma$ in Eq. (13) can be expressed as the double sum

$$H_e = \sum_\mu \sum_{\gamma \in G_\mu} q_\gamma \Phi_{\text{ext}}(\mathbf{r}_\gamma), \tag{14}$$

which partitions H_e into a sum over contributions associated to the DFT atoms μ. The proximity of the grid points $\gamma \in G_\mu$ to the positions $\mathbf{r}_\mu$ of the DFT atoms can now be exploited for the rapid evaluation of Eq. (14) by taking advantage of the SAMM$_p$ algorithm.[47]

C. Efficient computation of Φ_{ext}

According to the SAMM scheme,[34] the electrostatic potential at points $\mathbf{r}_\gamma$ in the vicinity of a given atom μ is calculated as a sum

$$\Phi_{\text{ext}}(\mathbf{r}_\gamma) = \sum_{l=0}^{l_{\max}} \Phi\left(\mathbf{r}_\gamma \mid C_\mu^l\right), \qquad \gamma \in G_\mu \tag{15}$$

over contributions $\Phi(\mathbf{r}_\gamma \mid C_\mu^l)$ from sources located in the distance classes C_μ^l, $l = 0, \ldots, l_{\max}$, to which for periodic systems a reaction field contribution $\Phi(\mathbf{r}_\gamma \mid \text{RF})$ is added[60] (for notational simplicity it will be omitted in the subsequent discussion).

Figure 2 illustrates how the external potential $\Phi_{\text{ext}}(\mathbf{r}_\gamma)$ is calculated using SAMM$_p$ for two PMM water molecules belonging to the distance classes C_μ^0 and C_μ^1 of a DFT atom μ. Here, the solid gray arrow marks the computation of the potential generated by the electrostatic moments of the PMM atoms $i \in C_\mu^0$ through

$$\Phi\left(\mathbf{r}_\gamma \mid C_\mu^0\right) = \sum_{i \in C_\mu^0} [\Phi(\mathbf{r}_\gamma \mid q_i, \mathbf{r}_i, \sigma_i) + \Phi(\mathbf{r}_\gamma \mid \mathbf{p}_i, \mathbf{r}_i, \tilde{\sigma}_i)] \tag{16}$$

with the potentials

$$\Phi(\mathbf{r}_\gamma \mid q_i, \mathbf{r}_i, \sigma_i) = \frac{q_i \operatorname{erf}[r_{\gamma i}/(\sqrt{2}\,\sigma_i)]}{r_{\gamma i}} \tag{17}$$

of Gaussian partial charges q_i and

$$\Phi(\mathbf{r}_\gamma \mid \mathbf{p}_i, \mathbf{r}_i, \tilde{\sigma}_i) = -\mathbf{p}_i \cdot \partial_{(1)} \Phi(\mathbf{r}_\gamma \mid q_i, \mathbf{r}_i, \tilde{\sigma}_i)/q_i \tag{18}$$

of Gaussian dipoles $\mathbf{p}_i^G$. In Eq. (18), the gradient is written as $\partial_{(1)}$.

For electrostatic PMM moments in all higher ($l \geq 1$) distance classes C_μ^l, the potentials

$$\Phi\left(\mathbf{r}_\gamma \mid C_\mu^l\right) = \sum_{n=0}^{p} \frac{1}{n!} \mathbf{r}_{\gamma\mu}^{(n)} \odot \mathbf{T}^{n,p}\left(\mathbf{r}_\mu \mid C_\mu^l\right) \tag{19}$$

are calculated through pth order Taylor expansions around the position $\mathbf{r}_\mu$ of the DFT atom μ. The symbol $\mathbf{r}_{\gamma\mu}^{(n)}$ is the n-fold outer product of $\mathbf{r}_{\gamma\mu}$ with itself. This vector $\mathbf{r}_{\gamma\mu}$ connects the reference point μ with the grid point $\gamma \in G_\mu$ (dotted gray arrow in Fig. 2). The symbol $\odot$ denotes the inner contraction product of two tensors (Ref. 47 thoroughly explains the employed tensorial notation). Finally, the class specific expansion coefficients

$$\mathbf{T}^{n,p}\left(\mathbf{r}_\mu \mid C_\mu^l\right) \equiv \partial_{(n)} \Phi^{\mathrm{T}}\left(\mathbf{r} \mid C_\mu^l\right)\big|_{\mathbf{r}_\mu} \tag{20}$$

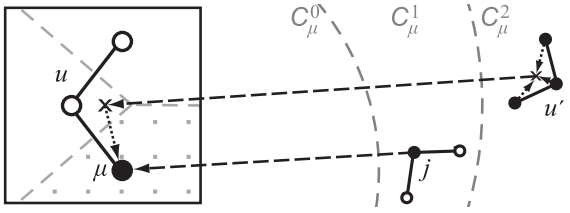


FIG. 3. Computation of Taylor expansion coefficients: The charges and induced dipoles of the PMM atoms $j \in C_\mu^1$ generate the coefficients $\mathbf{T}^{n,p}(\mathbf{r}_\mu \mid C_\mu^1)$ (lower dashed arrow). The PMM atoms in C_μ^2 are collected into structural units, whose electrostatic signatures are represented by multipole expansions.[47] For the PMM unit u', for instance, such an expansion is symbolized by three black dotted arrows pointing toward its reference point "×". The multipole potentials originating from u' are expanded into a Taylor series at the reference point of the DFT unit u (upper dashed arrow), from which the additional contributions $\mathbf{T}^{n,p}(\mathbf{r}_\mu \mid C_\mu^2)$ to the atom-centered expansion coefficients are inherited (dotted arrow) by a simple shifting procedure.[47]

are nth rank tensors generated by the nth order partial derivatives of the potentials $\Phi^{\mathrm{T}}(\mathbf{r} \mid C_\mu^l)$ at $\mathbf{r}_\mu$, which originate from point-like electrostatic moments (cf. the discussion of Fig. 1 above) occupying the distance class C_μ^l.

Here, the nature of the electrostatic moments, which generate the potential appearing in Eq. (20), differs for the distance classes at level $l = 1$ and at levels $l \geq 2$, respectively. As is schematically indicated by the lower dashed arrow in Figure 3, at $l = 1$ the point charges q_j and induced point dipoles $\mathbf{p}_j$ of the individual atoms j are considered to be the sources of the potential. The associated expansion coefficients

$$\mathbf{T}^{n,p}\left(\mathbf{r}_\mu \,\middle|\, C_\mu^1\right) = \partial_{(n)} \sum_{j \in C_\mu^1} [q_j - (1 - \delta_{np})\, \mathbf{p}_j \cdot \partial_{(1)}] \frac{1}{r_{\mu j}} \tag{21}$$

are essentially given by the nth rank tensors[47] $\partial_{(n)}(1/r)$. Here, the prefactor $1 - \delta_{np}$, in which δ is the Kronecker symbol, ensures that the expansion is of comparable accuracy for the atomic charges and dipoles.

For distance class levels $l \geq 2$, by contrast, mth order multipole moments ($m = 0, \ldots, p$) of nested and increasingly larger charge and dipole distributions are considered as the sources of the potential in Eq. (20). Within the SAMM$_p$ algorithm[47] the Taylor expansion coefficients

$$\mathbf{T}^{n,p}\left(\mathbf{r}_\mu \,\middle|\, C_\mu^l\right) = \partial_{(n)} \sum_{m=0}^{p-n} \Phi^m\left(\mathbf{r}_\mu \,\middle|\, C_\mu^l\right), \qquad n = 0, \ldots, p, \quad l \geq 2 \tag{22}$$

are then the partial derivatives of the mth order multipole potentials $\Phi^m(\mathbf{r}_\mu \mid C_\mu^l)$. As documented in the Appendix, the SAMM$_p$ algorithm, which has been originally developed for distributions of static charges,[47] can meanwhile also account for induced dipoles.

Figure 3 indicates for three atoms $k \in C_\mu^2$ collected into a unit u' the computation of multipole moments by dotted arrows and the calculation of the Taylor coefficients (22) by a two-step process (upper dashed and left dotted arrows) as is common in FMM methods.[47]

D. Forces on the DFT atoms μ

Equations (16) and (19) specify the two basically different procedures by which the external potential is imported onto the DFT grid. This import enables the DFT program CPMD to compute a polarized electron density $\rho_{\mathrm{e}}(\mathbf{r}_\gamma)$. Using the updated grid charges q_γ [cf. Eq. (12)] and the Gaussian core charge q_μ centered at $\mathbf{r}_\mu$, CPMD then calculates the electrostatic interaction energies

$$H_\kappa = \sum_\mu \sum_{l=0}^{l_{\max}} H_\kappa\left(C_\mu^l\right), \qquad \kappa \in \{\mathrm{e}, \mathrm{c}\}, \tag{23}$$

because of Eqs. (11) and (13)–(15) as sums of electronic

$$H_{\mathrm{e}}\left(C_\mu^l\right) = \sum_{\gamma \in G_\mu} q_\gamma \Phi\left(\mathbf{r}_\gamma \,\middle|\, C_\mu^l\right) \tag{24}$$

and nuclear

$$H_{\mathrm{c}}\left(C_\mu^l\right) = q_\mu \langle \Phi\left(\mathbf{r}_\mu \,\middle|\, C_\mu^l\right) \rangle_{\sigma_\mu} \tag{25}$$

contributions associated to the DFT atoms μ and distance class levels l. According to Eq. (9) an update of the interaction Hamiltonian $H_{\mathrm{DFT/(P)MM}}^{\mathrm{elec}}$ has thus been determined. By taking gradients of $H_{\mathrm{DFT}} + H_{\mathrm{DFT/(P)MM}}^{\mathrm{elec}}$ with respect to the coordinates $\mathbf{r}_\mu$, CPMD can now compute the electrostatic forces on the DFT atoms μ.

The electrostatic Hellmann-Feynman reaction forces[66] exerted by the charges q_γ and q_μ on the PMM atoms i follow from the gradients of $H_{\mathrm{DFT/(P)MM}}^{\mathrm{elec}}$ with respect to the coordinates $\mathbf{r}_i$. However, in the current form, the contributions (24) and (25) to the electrostatic interaction energy do not immediately reveal how $H_{\mathrm{DFT/(P)MM}}^{\mathrm{elec}}$ depends on the PMM coordinates $\mathbf{r}_i$ and, therefore, how these forces should be calculated. To uncover this dependence, the electronic and nuclear interaction energies $H_\kappa(C_\mu^l)$ will now be separately analyzed for the distance classes $l = 0$, 1, and $l \geq 2$.

E. Reaction forces on the PMM atoms $i \in C_\mu^0$

The electronic near-field interaction Hamiltonian

$$H_{\mathrm{e}}\left(C_\mu^0\right) = \sum_{\gamma \in G_\mu} q_\gamma \sum_{i \in C_\mu^0} [\Phi(\mathbf{r}_\gamma \mid q_i, \mathbf{r}_i, \sigma_i) + \Phi(\mathbf{r}_\gamma \mid \mathbf{p}_i, \mathbf{r}_i, \tilde{\sigma}_i)] \tag{26}$$

is obtained by inserting the external potential (16) originating from class C_μ^0 into Eq. (24). It is the energy of the point-like grid charges q_γ in the potentials of the Gaussian charges q_i of widths σ_i and Gaussian induced dipoles $\mathbf{p}_i$ of widths $\tilde{\sigma}_i$ collected in C_μ^0. Equations (17) and (18), respectively, specify these potentials in terms of the connection vectors $\mathbf{r}_{\gamma i}$ pointing from PMM atom i to the grid point γ.

Replacing these vectors by their inverses $\mathbf{r}_{i\gamma} = -\mathbf{r}_{\gamma i}$ and interchanging in Eq. (26), the sums over γ and i lead to the

strictly equivalent expression

$$H_e(C_\mu^0) = \sum_{i\in C_\mu^0}\Bigg[q_i \sum_{\gamma\in G_\mu}\Phi(\mathbf{r}_i \mid q_\gamma, \mathbf{r}_\gamma, \sigma_i) - \mathbf{p}_i \cdot \sum_{\gamma\in G_\mu}\mathbf{E}(\mathbf{r}_i \mid q_\gamma, \mathbf{r}_\gamma, \tilde{\sigma}_i)\Bigg], \tag{27}$$

which is the energy of point charges q_i and dipoles $\mathbf{p}_i$ at positions $\mathbf{r}_i$ in the potentials $\Phi(\mathbf{r}_i \mid q_\gamma, \mathbf{r}_\gamma, \sigma_i)$ and fields $\mathbf{E}(\mathbf{r}_i \mid q_\gamma, \mathbf{r}_\gamma, \tilde{\sigma}_i) = -\partial_{(1)}\Phi(\mathbf{r}_i \mid q_\gamma, \mathbf{r}_\gamma, \tilde{\sigma}_i)$ of Gaussian grid charges q_γ of widths σ_i and $\tilde{\sigma}_i$, respectively. The negative gradients of $H_e(C_\mu^0)$ with respect to the positions $\mathbf{r}_i$ are then the electrostatic Hellmann-Feynman forces[66] exerted by the grid charges q_γ, $\gamma \in G_\mu$, on the PMM atoms $i \in C_\mu^0$. If one inverts the solid gray arrow in Fig. 2, the inverted arrow can serve to symbolize such a reverse action of one of the grid charges on a nearby PMM atom.

The core contribution

$$H_c(C_\mu^0) = \sum_{i\in C_\mu^0}[q_i\Phi(\mathbf{r}_i \mid q_\mu, \mathbf{r}_\mu, \sigma_{\mu i}) - \mathbf{p}_i \cdot \mathbf{E}(\mathbf{r}_i \mid q_\mu, \mathbf{r}_\mu, \tilde{\sigma}_{\mu i})] \tag{28}$$

is analogously obtained by inserting the potential $\Phi(\mathbf{r}_\mu \mid C_\mu^0)$ as defined by Eq. (16) into Eq. (25), by executing the Gaussian averages (4) through an increase of the Gaussian widths (as explained in connection with Eq. (11)), and by repeating the arguments, which lead from Eq. (26) to Eq. (27).

F. Reaction forces on the PMM atoms $j \in C_\mu^1$

Inserting the Taylor expansion (19) with the coefficients (21) into (24) yields the electronic interaction energy

$$H_e(C_\mu^1) = \sum_{\gamma\in G_\mu} q_\gamma \sum_{n=0}^{p}\frac{1}{n!}\mathbf{r}_{\gamma\mu}^{(n)} \odot \partial_{(n)} \sum_{j\in C_\mu^1}[q_j - (1-\delta_{np})\,\mathbf{p}_j \cdot \partial_{(1)}]\frac{1}{r_{\mu j}} \tag{29}$$

for DFT atom μ and class C_μ^1. Interchanging the sums over j and γ and employing the identity $\partial_{(n)}(1/r_{\mu j}) = (-1)^n\partial_{(n)}(1/r_{j\mu})$, one finds

$$H_e(C_\mu^1) = \sum_{j\in C_\mu^1}\sum_{n=0}^{p}[q_j + (1-\delta_{np})\,\mathbf{p}_j \cdot \partial_{(1)}]\Phi^n(\mathbf{r}_j \mid Q_\mu, \mathbf{r}_\mu) \tag{30}$$

with the potentials

$$\Phi^n(\mathbf{r}_j \mid Q_\mu, \mathbf{r}_\mu) = \frac{(-1)^n}{n!}\left(\partial_{(n)}\frac{1}{r_{j\mu}}\right) \odot \sum_{\gamma\in G_\mu} q_\gamma \mathbf{r}_{\gamma\mu}^{(n)} \tag{31}$$

generated at the positions $\mathbf{r}_j$ of the PMM atoms by the nth order multipole moments of the electronic grid charges $Q_\mu \equiv \{q_\gamma \mid \gamma \in G_\mu\}$ of the DFT atom μ. According to Lorenzen *et al.*,[47] the potentials (31) can be equivalently written as

$$\Phi^n(\mathbf{r}_j \mid Q_\mu, \mathbf{r}_\mu) = \frac{(-2)^n}{(2n)!}\left(\partial_{(n)}\frac{1}{r_{j\mu}}\right) \odot \mathbf{M}^n(\mathbf{r}_\mu \mid Q_\mu). \tag{32}$$

Here,

$$\mathbf{M}^n(\mathbf{r}_\mu \mid Q_\mu) = \sum_{\gamma\in G_\mu} q_\gamma r_{\mu\gamma}^{2n+1}\left(\partial_{(n)}\frac{1}{r_{\mu\gamma}}\right) \tag{33}$$

are the reduced totally symmetric multipole tensors, which have only $2n+1$ independent components, because they are traceless with respect to every pair of tensor components.[67,68]

A slight rearrangement of Eq. (30) and the introduction of the multipole fields

$$\mathbf{E}^n(\mathbf{r}_j \mid Q_\mu, \mathbf{r}_\mu) = -\partial_{(1)}\Phi^n(\mathbf{r}_j \mid Q_\mu, \mathbf{r}_\mu) \tag{34}$$

finally leads to

$$H_e(C_\mu^1) = \sum_{j\in C_\mu^1}\Bigg[q_j \sum_{n=0}^{p}\Phi^n(\mathbf{r}_j \mid Q_\mu, \mathbf{r}_\mu) - \mathbf{p}_j \cdot \sum_{n=0}^{p-1}\mathbf{E}^n(\mathbf{r}_j \mid Q_\mu, \mathbf{r}_\mu)\Bigg], \tag{35}$$

which explicitly reveals the desired dependence on the PMM coordinates $\mathbf{r}_j$ and, therefore, enables the derivation of likewise simple expressions for the Hellmann-Feynman forces on the PMM atoms $j \in C_\mu^1$.

In the contributions (25) to the core Hamiltonian, the Gaussian average can be neglected at all levels $l \geq 1$. At the reference point $\mathbf{r}_\mu$, i.e., for $\mathbf{r}_{\gamma\mu} = \mathbf{0}$, the Taylor expansion (19) reduces to the zeroth order term $\mathbf{T}^{0,p}(\mathbf{r}_\mu \mid C_\mu^l)$. Inserting this result into Eq. (25) yields the general form

$$H_c(C_\mu^l) = q_\mu \mathbf{T}^{0,p}(\mathbf{r}_\mu \mid C_\mu^l) \quad \text{for} \quad l \geq 1. \tag{36}$$

Inserting for $l = 1$ the coefficients (21) into (36) and repeating the steps, which lead from (29) to (35), one gets

$$H_c(C_\mu^1) = \sum_{j\in C_\mu^1}[q_j\Phi(\mathbf{r}_j \mid q_\mu, \mathbf{r}_\mu) - \mathbf{p}_j \cdot \mathbf{E}(\mathbf{r}_j \mid q_\mu, \mathbf{r}_\mu)], \tag{37}$$

which is the energy of the PMM atoms $j \in C_\mu^1$ in the potential and field of the point-like core charge q_μ. Instead of separately evaluating Eq. (37), one may equivalently include the core charge q_μ into the grid charge distribution Q_μ of DFT atom μ, which then becomes $\hat{Q}_\mu = Q_\mu \cup q_\mu$. Because q_μ is located by construction at the reference point of the multipole expansion of $\hat{Q}_\mu$, its inclusion solely modifies the monopole moment to $\mathbf{M}^0(\mathbf{r}_\mu \mid \hat{Q}_\mu) = \mathbf{M}^0(\mathbf{r}_\mu \mid Q_\mu) + q_\mu$.

G. Reaction forces on the PMM atoms $k \in C_\mu^l$, $l \geq 2$

Starting at level $l = 2$, the computation of the electrostatic interactions becomes identical to the SAMM$_p$ treatment of a purely classical system, which has been described in detail elsewhere.[47] Therefore, it suffices here to sketch how at level $l = 2$ a DFT fragment is integrated into the SAMM$_p$ algorithm.

For a most simple presentation, we assume that the DFT fragment is composed of a single structural unit u like in the example depicted in the above figures. All its atoms μ share a common distance class C_u^2 ($\forall\mu \in u : C_\mu^2 = C_u^2$), which contains the distant PMM atoms interacting on level $l = 2$ with u. Equations (19) and (22)–(24) yield the associated electronic

interaction energy

$$H_e(C_u^2) = \sum_{\mu\in u}\sum_{\gamma\in G_\mu} q_\gamma \sum_{n=0}^{p} \frac{1}{n!}\mathbf{r}_{\gamma\mu}^{(n)} \odot \partial_{(n)} \sum_{m=0}^{p-n} \Phi^m(\mathbf{r}_\mu \mid C_u^2). \tag{38}$$

In this formulation, the potentials $\Phi^m(\mathbf{r}_\mu \mid C_u^2)$ generated by mth order multipole moments of the PMM structural units $u' \in C_u^2$ are considered as sources and the grid charges of the DFT unit u as targets of the electrostatic interactions. These targets are addressed through Taylor expansions around the atomic positions $\mathbf{r}_\mu$.

SAMM$_p$ does not evaluate the nth order expansion coefficients $\partial_{(n)} \sum_{m=0}^{p-n} \Phi^m(\mathbf{r}_\mu \mid C_u^2)$ at each $\mathbf{r}_\mu \in u$, but only at the reference point $\mathbf{r}_u$ (left "×" in Fig. 3) of unit u. Using a Taylor expansion around $\mathbf{r}_u$, they are then simply shifted to the atomic positions $\mathbf{r}_\mu$.[47] By the very construction of SAMM$_p$, this shifting is exactly the inverse operation to the combination[47] of all atomic nth order multipole moments $\mathbf{M}^n(\mathbf{r}_\mu \mid Q_\mu)$, $\mu \in u$, which are known as soon as the energies $H_e(C_\mu^1)$ have been calculated through Eq. (35), into corresponding moments $\mathbf{M}^n(\mathbf{r}_u \mid Q_u)$ of the charge distribution $Q_u \equiv \bigcup_{\mu\in u} Q_\mu$ of the DFT unit u. Note that this symmetry of Taylor and multipole expansions is the reason why in (P)MM simulations the reaction principle holds exactly for the SAMM$_p$ forces. In the given DFT/(P)MM case, however, the grid discretization of ρ_e weakly interferes with this principle through an artifact, which we will address further below.

Due to the quoted symmetry one can equivalently represent the interaction energy (38) in a form in which the multipole moments $\mathbf{M}^n(\mathbf{r}_u \mid Q_u)$ of the grid charge distribution Q_u are the sources of multipole potentials $\Phi^m(\mathbf{r} \mid Q_u)$ acting on distant PMM atoms $k \in C_u^2$ through local Taylor expansions. The Hellmann-Feynman forces on these PMM atoms immediately follow from the corresponding SAMM$_p$ expression (for details and explanations, see Ref. 47). We note that the interaction $H_c(C_u^2)$ of the atomic cores in unit u with the PMM atoms $k \in C_u^2$ is included, if one employs the extended atomic multipole moments $\mathbf{M}^n(\mathbf{r}_\mu \mid \hat{Q}_\mu)$ instead of the electronic moments $\mathbf{M}^n(\mathbf{r}_\mu \mid Q_\mu)$ for computing the moments of unit u.

The analysis given above for the case of a single DFT unit u interacting with PMM atoms in the distance classes C_u^2 is readily generalized to higher cluster levels and larger DFT fragments. In our implementation, the electrostatic DFT/(P)MM interactions are calculated at levels $l \geq 2$ by transferring the atomic multipole moments $\mathbf{M}^n(\mathbf{r}_\mu \mid \hat{Q}_\mu)$ computed by CPMD to the (P)MM-MD program IPHIGENIE, which then calculates the multipole moments $\mathbf{M}^n(\mathbf{r}_u \mid \hat{Q}_u)$ of unit u. From now on the moments of DFT units are treated at all SAMM$_p$ levels $l \geq 2$ just like (P)MM moments. The hierarchically nested FMM scheme then renders the total electrostatic forces on the PMM atoms $k \in C^l$, $l \geq 2$.

H. Remarks

The evaluation of $\sum_{l\geq 1} H_e(C_\mu^l)$ is computationally about as expensive as the evaluation of $H_e(C_\mu^0)$ for a single PMM atom i [cf. Eq. (27)]. Therefore, the computational effort spent on H_e is essentially determined by the average number $N_0 \equiv (1/N_{\mathrm{DFT}}) \sum_\mu |C_\mu^0|$ of PMM atoms found in the innermost distance classes C_μ^0 of the N_{DFT} DFT atoms μ. Typically one finds $N_0 \approx 100$ and, therefore, the computational advantage of the above calculation scheme over a brute force method is N_0/N, if N is the number of PMM atoms in the system. Thus, for a typical simulation system with $N = 10^4$, the speedup is about 10^2.

The computational scheme described in Sec. II F for the interactions of the DFT grid charges with the PMM atoms in distance class C_μ^1 resembles the DFT/MM suggestion of Laio *et al.*,[38] which also applies Taylor and multipole expansions on the DFT grid to compute interactions with distant MM atoms. These expansions are centered for the whole DFT grid around a single reference point, are truncated at the order $p = 2$, and treat all distant MM atoms as individual sources and targets of electrostatic interactions. Our approach, by contrast, partitions the DFT grid by N_{DFT} reference points, extends the local expansions up to order $p = 4$, and considers for each DFT atom only the comparably few PMM atoms, which are in the small distance range from about 6 Å to about 8 Å, as individual sources and targets of these expansions while collecting all more distant atoms into a hierarchy of increasingly large clusters. Thus, our approach should be much more accurate and efficient even for relatively small DFT/MM systems. Like the scheme of Laio *et al.*,[38] our approach also does not correct the small force discontinuities occurring whenever atoms change distance classes. However, in our case the effects of these transitions are smaller, because the forces are calculated with higher level multipole and Taylor expansions.

III. KEY POINTS OF THE IMPLEMENTATION

Section II completely covers the basic theory of our DFT/PMM approach. However, for an energy conserving and computationally efficient implementation two important issues must be additionally considered.

A. Movements of the DFT box

In DFT/(P)MM dynamics simulations, the grid-based representation of ρ_e by CPMD interferes with energy conservation. In CPMD, the energy E of a DFT atom μ depends on its relative position within the grid. Shifting, e.g., its position $\mathbf{r}_\mu$ along the line connecting a grid point γ with one of its nearest neighbors entails a sinusoidal modulation $E(|\mathbf{r}_\mu - \mathbf{r}_\gamma|) \sim -\Delta E \cos(2\pi |\mathbf{r}_\mu - \mathbf{r}_\gamma|/a)$, where a is the associated grid constant. In the DFT setting applied by us (see Sec. IV), the relative modulation $\Delta E/E(0)$ is about 10^{-5}%. Thus, the atom prefers to sit at grid points and experiences artificial grid forces at other positions. As long as the grid remains fixed in space, this small grid artifact solely represents a rough background potential, whose contributions to the total energy on average vanish during a dynamics simulation.

However, if the DFT fragment moves during a DFT/(P)MM dynamics simulation, the DFT box has to follow. Such a movement of the discretized box may lead to random forces on the DFT atoms adding heat to the system.

This serious artifact can be avoided, if the DFT grid is considered as an infinite object, on which the DFT box is shifted in units of the lattice constants whenever the movement of the DFT fragment (as measured, e.g., by its center of geometry) exceeds the lattice constant in one of the three spatial directions. Thus, only those box translations are allowed, which would leave a fully periodic DFT system invariant.

B. DFT/PMM-SCF iteration

The polarizable degrees of freedom of a DFT/PMM system, i.e., the PMM dipole strengths $\mathbf{p}_j$ and the DFT electron density ρ_e, have to be calculated in coupled SCF procedures, which can be rapidly brought to convergence by diligent choices of the initial conditions. Assuming that the $\mathbf{p}_j$ and the Kohn-Sham orbitals determining ρ_e are temporally continuous during the integration of the dynamics, these entities can be extrapolated from a history of $M^h \approx 4$ previous integration steps using Lagrangian polynomials.[69] During the PMM-SCF iteration, the "direct inversion of the iterative subspace" (DIIS) algorithm[70] with a history length $\hat{M}^{h,p} \approx 3$ is used to speed up convergence. Similarly, CPMD[48] applies DIIS[71] during DFT-SCF with $\hat{M}^{h,\rho} = 10$.

After the integration of the nuclear motion, the potential Φ_{ext} polarizing the DFT fragment is computed from the static partial charges and from the extrapolated dipole strengths $\mathbf{p}_j^0 = \mathbf{p}_j^h$ in the PMM fragment. Next, the DFT-SCF iteration is executed with a loose initial convergence criterion $\chi_{DFT}^{ini} = 10\,\chi_{DFT}$, which limits the largest element of the gradient of the wave function.[48] Keeping the resulting first guess ρ_e fixed, the $\mathbf{p}_j$ are iterated until in the nth step $|\mathbf{p}_j^{n-1} - \mathbf{p}_j^n| < \chi_{PMM}$ for all j, where χ_{PMM} is a certain threshold. The resulting $\mathbf{p}_j$ modify the external potential, to which the DFT fragment is exposed. In subsequent DFT-SCF calculations, the tight DFT-SCF criterion χ_{DFT} is used if the preceding PMM-SCF iteration converged within one iteration step or if a predefined number k of DFT-SCF calculations is exceeded. Thus, in the default case $k = 1$, only the first DFT-SCF calculation after an integration step is performed with χ_{DFT}^{ini}.

We will show that this strategy avoids extended and costly DFT-SCF iterations as long as the $\mathbf{p}_j$ are far from convergence. It partially resembles a scheme proposed by Thompson and Schenter[8,9] in the context of QM/PMM, where QM stands for semi-empirical quantum chemistry.

IV. METHODS

For the examination of our new DFT/PMM method, we employed two different simulation systems, a water dimer and a periodic box of liquid PMM water containing one DFT water molecule. The dynamics was integrated by the Verlet algorithm[72] with a time step $\Delta t = 0.25$ fs for the dimer and $\Delta t = 0.5$ fs for the DFT/PMM liquid. The geometries of the respective PMM water models were fixed using MSHAKE[73] and the electrostatics was treated at the SAMM$_4$ level.[47] The respective DFT water molecule was described by the gradient-corrected exchange functional of Becke[74] together with the correlation functional of Perdew (BP),[75] and the norm-conserving pseudo-potentials of Troullier and Martins (MT).[76] It was centered into a cubic box of edge length 9 Å containing the grid of the plane wave basis set, which was cutoff at 70 Ry. We denote this particular DFT approach by MT/BP.

A. Water dimer

The energy conservation was checked by 2 ps MD simulations of water dimers. Here the initial velocities indicated a temperature of about 80 K. We adopted the SCF convergence criteria $\chi_{PMM} = 10^{-4}$ D and $\chi_{DFT} = 10^{-7}$. In the DFT/PMM hybrid setting, the H-bond donor was described by MT/BP and the acceptor by the initial version TL4P$_{ini}$ of a recently developed PMM water model (Tröster *et al.*[90]). To provide references, we simulated a MT/BP dimer using a (15 Å)^3 DFT box and a TL4P$_{ini}$ dimer.

TL4P$_{ini}$ features the experimental liquid phase geometry[77,78] ($l_{OH} = 0.968$ Å, $\varphi_{HOH} = 105.3°$), the experimental gas phase dipole moment[79] (1.85 D) and polarizability[80] (1.47 Å^3), a massless negative charge $q_M = -1.172$ e on the bisectrix 0.258 Å distant from the oxygen, and positive charges at the hydrogens as well as a Gaussian inducible dipole of width $\tilde{\sigma}_i = 0.912$ Å at the oxygen. The van der Waals interactions were treated identically for all components of the hybrid systems, i.e., by Buckingham potentials[81] $E_B(r) = A_1 \exp(-rA_2) - B/r^6$ centered at the oxygen atoms ($A_1 = 78700$ Å^{12} kcal/mol, $A_2 = 3.50$ Å^{-1}, $B = 1062$ Å^6 kcal/mol). For the Gaussian distributions, which represent the static partial charges of the PMM atoms as long as they are close to the DFT atoms, we chose identical widths $\sigma_i = 0.57$ Å as suggested in Ref. 34.

In the unrestrained hybrid dimer, the PMM fragment moves close to the DFT fragment thus probing the innermost distance classes at the level $l = 0$. To check the electrostatics treatment also for outer distance classes (levels $l = 1, 2$), we softly restrained the distance d_{OO} of the two oxygens by a harmonic potential with a spring constant of 1 kcal/mol Å^2 to $d_{OO,1} = 7$ Å and $d_{OO,2} = 10$ Å, respectively, thereby guaranteeing that the interactions were calculated within the distance class level $l \in \{1, 2\}$ of interest. This probing of outer distance classes was also used in reference simulations of the PMM dimer.

B. Liquid water

The stability and performance of the DFT/PMM algorithm were investigated with a periodic cubic box [volume $V = (46.6$ Å$)^3$] filled with $N = 3374$ TL4P$_{ini}$ water molecules. Thus, the experimental density[82] $n = 0.9965$ g/cm^3 of water at the temperature $T = 300$ K and the pressure $p = 1$ atm was prepared. Extending the SAMM$_4$ treatment of the electrostatics by a moving boundary reaction field correction[60] and modeling the surrounding dielectric continuum by a dielectric constant of 80, the system was equilibrated for 1 ns in the NVT ensemble. Here, T was kept at 300 K with a Bussi thermostat[83] (relaxation time 0.1 ps).

To check the long-time stability, ten snapshots were drawn from the last 100 ps of this trajectory. Each snapshot

served as a starting point for a 14 ps DFT/PMM-MD simulation, in which one of the water molecules was described by MT/BP and the thermostat was restricted to the PMM environment.

Starting a series of 250 fs DFT/PMM-MD simulations at identical initial conditions, we studied how the efficiency and accuracy of the algorithm are affected by the SCF convergence criteria χ_{PMM} and χ_{DFT}. The accuracy was assessed by comparing the temporal evolutions of the DFT fragment's energy $E(t)$ and dipole moment $\mathbf{p}(t)$. Replacing the polarizable TL4P$_{\mathrm{ini}}$ water models by non-polarizable TIP4P/2005 potentials[84] and equilibrating this MM system like its PMM counterpart enabled a DFT/MM reference simulation.

V. TEST SIMULATIONS

Numerical integrations of the Hamiltonian dynamics employ time steps Δt of finite size. Therefore, the total energy $E(t)$ shows small fluctuations $\Delta E(t \mid \Delta t)$ around an average value $\langle E \rangle(\Delta t)$.[85] In the limit $\Delta t \to 0$, the fluctuations vanish to leading order with Δt^2. Similarly the average converges to E_0, which is the conserved value of the true Hamiltonian.

Thus, one expects small fluctuations $\Delta E(t \mid \Delta t)$ of the total energy $E(t)$ around its constant average $\langle E \rangle(\Delta t)$ also for the reference dynamics simulation of a PMM water dimer, in which all forces are calculated as exact negative gradients of the potential energy. Figure 4(a) demonstrates that this is actually the case for the unrestrained PMM water dimer at close contact. Here the standard deviation $\sigma_{\Delta E}$ of $\Delta E(t \mid \Delta t)$ is 20×10^{-6} kcal/mol. According to Fig. 4(b), the energy is likewise well-conserved for the PMM dimer softly restrained to distances, which are large enough to enable the approximate treatment of the electrostatics by the SAMM$_4$ algorithm. Here, the standard deviation $\sigma_{\Delta E}$ is only 0.4×10^{-6} kcal/mol, because the forces are smaller by more than one order of magnitude. As expected,[85] for both distance classes the standard deviations $\sigma_{\Delta E}$ vanish and the averages $\langle E \rangle(\Delta t)$ converge with Δt^2. Section S1 of the supplementary material[86] provides evidence for these claims.

Figure 4(c) shows for the reference Born-Oppenheimer (BO) dynamics of the unrestrained DFT dimer the deviation $\Delta E(t \mid \Delta t)$ of the total energy $E(t)$ from its average $\langle E \rangle(\Delta t)$. It exhibits much larger fluctuations $\Delta E(t \mid \Delta t)$ than its PMM relative in Fig. 4(a) as quantified by the standard deviation $\sigma_{\Delta E} = 1.6 \times 10^{-3}$ kcal/mol. Also these fluctuations vanish with Δt^2 and their enhanced magnitude can be largely attributed to the high frequency O–H stretching modes in the DFT dimer, which are absent in the constrained PMM dimer. Fluctuations of a comparable size have been previously reported for the DFT simulation of a water trimer[39] with the related grid code CP2K.[87]

A. Energy conservation in DFT/PMM-MD

Figure 5(a) shows the energy fluctuations observed for the unrestrained DFT/PMM hybrid dimer at close contact. A visual comparison with Fig. 4(c) immediately demonstrates that the average DFT/PMM energy is as well conserved as for the DFT reference and that the DFT/PMM energy fluctuations are of comparable size.

The conservation of the average energy is also observed in Figs. 5(b) and 5(c), which pertain to the DFT/PMM dimers restrained at distances $d_{\mathrm{OO}} \approx 7$ Å and $d_{\mathrm{OO}} \approx 10$ Å. Here the standard deviation $\sigma_{\Delta E}$ of the energy fluctuations is a little larger than at close contact. The similarity of the DFT/PMM fluctuations $\Delta E(t \mid \Delta t)$ to those of the DFT reference becomes even more striking, if one studies the graphs in Figs. 4(c) and 5 at a higher time resolution. A corresponding graphical illustration is provided by Figure S11 in Sec. S2 of the supplementary material.[86]

The above data lead to the conclusion that our DFT/PMM interaction scheme conserves the energy of a hybrid system at all distance classes, because the sample dimers were studied

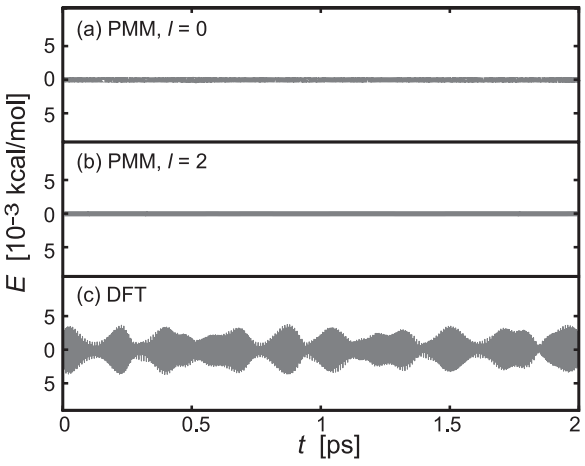


FIG. 4. Energy conservation in reference simulations of the water dimer. (a) PMM dynamics at close contact (exact electrostatics, $l = 0$) and (b) softly restrained to a distance $d_{\mathrm{OO}} \approx 10$ Å (SAMM$_4$ electrostatics, $l = 2$). (c) DFT Born-Oppenheimer dynamics at close contact.

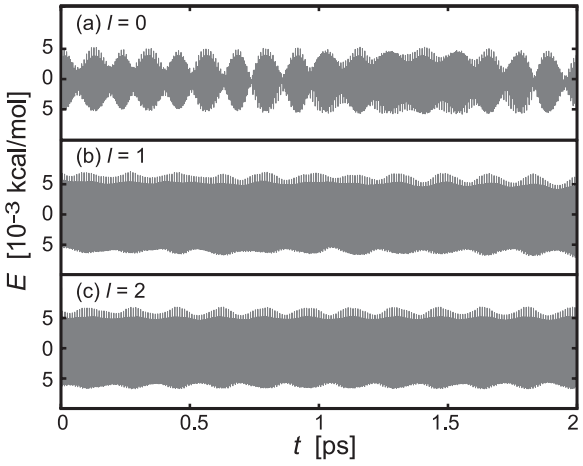


FIG. 5. Energy conservation in DFT/PMM hybrid simulations of the water dimer with the electrostatics treated at different distance class levels l. (a) Close contact ($l = 0$), (b) softy restrained to $d_{\mathrm{OO}} \approx 7$ Å ($l = 1$), and (c) to $d_{\mathrm{OO}} \approx 10$ Å ($l = 2$).

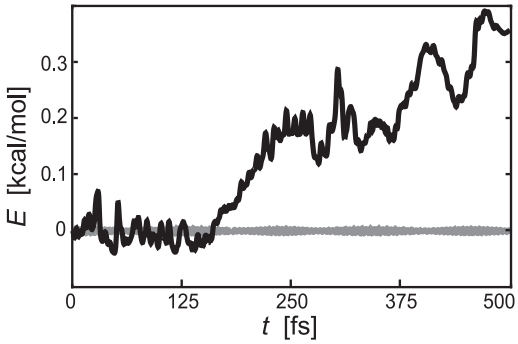


FIG. 6. Influence of the algorithm for DFT box movement on the energy conservation as exemplified by the DFT/PMM water dimer at close contact. The black and gray lines show the trajectories of the total energy in a simulation using a naive and our refined algorithm, respectively (see the text for explanation). The gray line represents the data of Fig. 5(a) on a different energy scale.

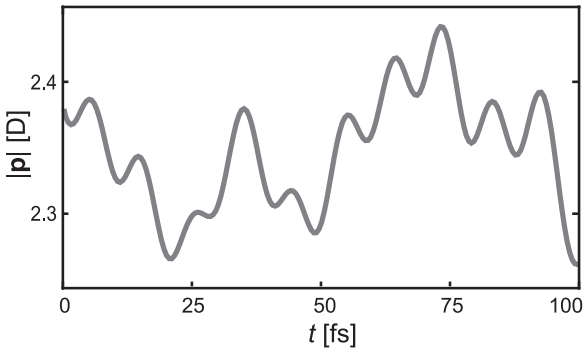


FIG. 7. The absolute value $|\mathbf{p}(t)|$ of the DFT fragment's dipole moment during a MD simulation of the aqueous DFT/PMM system described in Sec. IV. A short (100 fs) section of a trajectory was chosen to visualize the fluctuations of $|\mathbf{p}(t)|$ at a high time resolution.

at the interaction levels $l = 0$, 1, and 2, which cover all relevant algorithmic features discussed above. Note here that the energy was equally well conserved in DFT/MM simulations of a water dimer, in which the TIP3P potential[45] was applied to the MM fragment (data not shown).

Using the DFT/PMM water dimer at close contact as an example, we now additionally demonstrate with Figure 6 how the box movement algorithm described in Sec. III A supports energy conservation. The algorithm adapts the position of the DFT box to the motion of the DFT fragment only occasionally by using discrete displacements, which leave the (infinitely extended) DFT grid invariant.

The gray line in Fig. 6 shows the trajectory $E(t)$ of the total dimer energy for our grid-commensurate algorithm of box movement. The black line is an alternative trajectory, which was obtained by naively moving the DFT box at every MD time step with the center of geometry of the DFT fragment. In this case serious algorithmic artifacts apparently hamper energy conservation.

B. Smoothness and stability of DFT/PMM-MD

The ten 14 ps DFT/PMM-MD simulations of the periodic liquid water box described in Sec. IV clearly revealed the long-time stability of the algorithm. The calculated trajectories turned out to be smooth and did not show any artifacts.

Figure 7 exemplifies this smoothness at an elevated time resolution for the absolute value $|\mathbf{p}|$ of the dipole moment, which was calculated for the DFT fragment. The depicted 100 fs section represents an arbitrary choice from one of the 14 ps trajectories. The observed fluctuations of $|\mathbf{p}(t)|$ are caused by the thermal motions of the DFT molecule and of its surrounding $TL4P_{ini}$ counterparts. Because of their smoothness, one can calculate condensed phase IR spectra of DFT solute molecules from such DFT/PMM trajectories using Fourier transform methods.[36,88]

C. Performance of DFT/PMM-MD

Taking the DFT/PMM liquid water system as an example, Figure 8 gives an overview over the performance of our DFT/PMM algorithm for the choices $\chi_{PMM} = 10^{-4}$ D and $\chi_{DFT} = 10^{-6}$ of the SCF convergence parameters. For this check, CPMD was run in an MPI (message passing interface) parallel version[89] using eight 1.86 GHz cores distributed on two Intel Xeon E5320 processors while IPHIGENIE was executed in sequential mode on one of these cores.

As shown by a sample simulation, the integration of a periodic PMM system of the given size, which exclusively consists of $TL4P_{ini}$ water models, is only by a factor of 4.2 slower than that of a MM system made up of TIP4P/2005 models.[84] For hybrid settings, the comparison of the first two bars in Fig. 8 reveals that T_{PMM} is a factor of six larger than T_{MM}. This increase of T_{PMM} is caused by the additional polarizing action of the DFT fragment on the induced PMM dipoles, which costs on average one additional PMM-SCF iteration step.

For our DFT/PMM water box, T_{PMM} is about as large as the average time T_{DFT} spent with eight processors on the DFT part (cf. Fig. 8). T_{DFT} is composed of the times T_{DFT}^{SCF}, T_{DFT}^{import}, and T_{DFT}^{export} spent for the DFT-SCF iterations, for the import of the electrostatics onto and for its export from the DFT grid, respectively. Figure 8 shows that T_{DFT}^{SCF} is the main contribution to T_{DFT}.

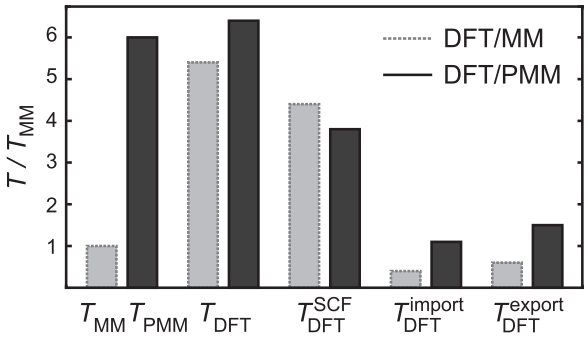


FIG. 8. Average computing times (walltimes) spent for our liquid water sample system per MD integration step on the various parts of a DFT/PMM (black) or DFT/MM (gray) calculation. Here, the DFT part was executed in parallel on eight core and the (P)MM part sequentially on one core. The time T_{MM} spent on the MM part in the DFT/MM setting is taken as the reference. For explanation see the text.

If χ_{DFT} is multiplied by ten, the number of DFT-SCF iterations and, hence, T_{DFT}^{SCF} are reduced by a factor of about 1/2. Fortunately, this increase of χ_{DFT} does not change the trajectories of the DFT fragment's energy $E(t)$ and dipole moment $\mathbf{p}(t)$ within the chosen 250 fs time window (Figure S12 in Sec. S3 in the supplementary material[86] provides a relevant example). Therefore, a criterion $\chi_{DFT} = 10^{-5}$ could be already sufficient for future DFT/(P)MM simulations.

Figure 8 indicates that T_{DFT}^{SCF} is by 15% smaller for DFT/PMM than for DFT/MM. This advantage of a PMM over a MM environment is apparently caused by our choice of an initially loose DFT-SCF convergence criterion $\chi_{DFT}^{ini} = 10\,\chi_{DFT}$ in the joint DFT/PMM-SCF iteration. If one applies only a single criterion ($\chi_{DFT}^{ini} = \chi_{DFT}$) instead, the time T_{DFT}^{SCF} of DFT/PMM is by 20% larger than that of DFT/MM (data not shown).

Despite the computational benefit caused by our two-stage DFT-SCF convergence criterion in the DFT/PMM setting, T_{DFT} is larger than in the DFT/MM case, because the times T_{DFT}^{import} and T_{DFT}^{export} used for the import and the export of the electrostatics are larger by factors 2.7 and 3.0, respectively. These factors mainly reflect the fact that for DFT/MM the DFT module is called only once per integration step, whereas in the DFT/PMM case it is called several times (in the given example: 2.0 DFT calls/integration step). Note that each such call requires the import and export of the electrostatics, which are more expensive, if induced dipoles are additionally present.

Finally, we examined the effect of varying the PMM convergence criterion. Tightening χ_{PMM} by one order of magnitude entails a slight increase of the numbers of PMM-SCF iteration steps, of DFT calls per integration step, and of DFT-SCF iteration steps. By contrast, loosening χ_{PMM} by one order of magnitude leads to a costly increase of the number of DFT-SCF iteration steps. Apparently, a correspondingly noisier PMM dipole environment hampers the DFT-SCF convergence.

In the DFT/PMM water box example discussed above, which features a small DFT fragment in a large PMM environment, the parallel DFT computation with eight cores requires about as much time as the single core PMM computation. This finding points to a technical deficiency of our current implementation, according to which the MPI-parallel version of the PMM-MD program IPHIGENIE is not yet compatible with the MPI-parallel version of the DFT program CPMD. This remaining deficiency will be removed in the near future.

VI. SUMMARY AND OUTLOOK

By suitably combining the (P)MM-MD program IPHIGENIE[47] with the DFT program CPMD,[48] we have developed and implemented a Hamiltonian DFT/(P)MM-MD approach, which conserves the energy as good as the well-established DFT-Born-Oppenheimer MD approach implemented in CPMD. Here, we took advantage of the fast multipole method $SAMM_4$ implemented in IPHIGENIE, by which one can treat the long-range electrostatic interactions within a (P)MM simulation system in a linearly scaling and Hamiltonian fashion. Artificial distortions of the DFT electron density are excluded by the use of Gaussian charges and induced dipoles in the boundary region between the DFT and PMM fragments. The accuracy and efficiency of the new DFT/PMM interface are supported by algorithmic improvements concerning the adaptive repositioning of the DFT box and the fine-tuning of the joint DFT- and PMM-SCF cycles. All these aspects were illustrated by relevant test simulations, which demonstrate that the new interface opens the way toward temporally extended DFT/PMM-MD simulations of large condensed phase systems at well-defined thermodynamic conditions.

Beyond the ongoing efforts of extending the implementation toward a jointly parallelized treatment of the DFT and (P)MM fragments, only one important issue remains to be addressed before applications can be tackled. It concerns the widths σ_i, which have to be chosen for the Gaussian distributions representing the (P)MM partial charges in the vicinity of the DFT fragment. As explained already by Laio *et al.*,[38] these widths are decisive parameters of the interface model and must be carefully determined for the various (P)MM atom types by sample simulations. Here, the value $\sigma_i = 0.57$ Å most likely is a sub-optimal choice.

Concerning the future development, we plan to reorganize the interface in such a way that also other grid-based DFT programs like CP2K[87] can be employed. Here, the integration of the multigrid DFT/MM electrostatic coupling of Laino *et al.*[39] into our near-field electrostatics computation should yield further efficiency gains. Furthermore, a combination of our DFT/PMM approach with the SCP-DFT method[32,33] implemented in CP2K may pave the way toward an improved modeling of polarization and dispersion interactions between the fragments.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (SFB749/C4) and by the Kompetenznetzwerk für wissenschaftliches Höchstleistungsrechnen in Bayern of the Bayerische Staatsministerium für Wissenschaft, Forschung und Kunst (KONWIHR-III).

APPENDIX: MULTIPOLE MOMENTS $\tilde{\mathbf{M}}^{m,\mathbf{0}}$ OF DIPOLE DISTRIBUTIONS

Compact formulas, by which the totally symmetric and traceless mth order multipole tensors $\mathbf{M}^{m,\mathbf{0}}$ can be calculated for $m = 0, 1, 2, \ldots$, if the origin $\mathbf{0}$ of the Cartesian coordinate system is chosen as the reference point, have been given in Sec. 3 of the supporting information of Ref. 47 for distributions B of partial charges.

Choosing the same setting, the corresponding tensors

$$\tilde{\mathbf{M}}^{m,\mathbf{0}} = \sum_{j\in\tilde{B}} \tilde{\mathbf{M}}_j^{m,\mathbf{0}} \tag{A1}$$

of rank $m = 2, \ldots$ can be calculated for a distribution $\tilde{B}$ of point dipoles $\mathbf{p}_j$ at positions $\mathbf{r}_j$ from the recursion

$$\tilde{\mathbf{M}}_j^{m,\mathbf{0}} = \hat{S}_m \left[\left(\frac{2m^2 - m}{m-1} \right) \left(\mathbf{r}_j \otimes \tilde{\mathbf{M}}_j^{m-1,\mathbf{0}} \right) - m \left(\mathbf{r}_j \odot \tilde{\mathbf{M}}_j^{m-1,\mathbf{0}} \right) \otimes \mathbf{I} \right], \tag{A2}$$

where $\hat{S}_n$ is the symmetrization operator given in Eq. (22) of Ref. 47, where also the applied tensorial notation is explained in detail. For dipole distributions, the monopole moment ($m = 0$) vanishes, of course, and the recursion (A2) is initialized with the point dipole

$$\tilde{\mathbf{M}}_j^{1,\mathbf{0}} = \mathbf{p}_j. \tag{A3}$$

For $m = 2, 3, 4$, the Cartesian components of the tensors $\tilde{\mathbf{M}}^{m,\mathbf{0}}$ are explicitly given by

$$\tilde{M}_{\alpha\beta}^{2,\mathbf{0}} = \sum_{j \in \tilde{B}} [3 r_{j\alpha} p_{j\beta} + 3 r_{j\beta} p_{j\alpha} - 2\delta_{\alpha\beta} (\mathbf{r}_j \cdot \mathbf{p}_j)], \tag{A4}$$

$$\begin{aligned} \tilde{M}_{\alpha\beta\gamma}^{3,\mathbf{0}} = 3 \sum_{j \in \tilde{B}} [& 5 (p_{j\alpha} r_{j\beta} r_{j\gamma} + p_{j\beta} r_{j\gamma} r_{j\alpha} + p_{j\gamma} r_{j\alpha} r_{j\beta}) \\ & - {r_j}^2 (p_{j\alpha} \delta_{\beta\gamma} + p_{j\beta} \delta_{\gamma\alpha} + p_{j\gamma} \delta_{\alpha\beta}) \\ & - 2 (\mathbf{r}_j \cdot \mathbf{p}_j) (r_{j\alpha} \delta_{\beta\gamma} + r_{j\beta} \delta_{\gamma\alpha} + r_{j\gamma} \delta_{\alpha\beta})], \end{aligned} \tag{A5}$$

$$\begin{aligned} \tilde{M}_{\alpha\beta\gamma\epsilon}^{4,\mathbf{0}} = 3 \sum_{j \in \tilde{B}} \{ & 35 (p_{j\alpha} r_{j\beta} r_{j\gamma} r_{j\epsilon} + p_{j\beta} r_{j\gamma} r_{j\epsilon} r_{j\alpha} \\ & + p_{j\gamma} r_{j\epsilon} r_{j\alpha} r_{j\beta} + p_{j\epsilon} r_{j\alpha} r_{j\beta} r_{j\gamma}) \\ & - 5 {r_j}^2 [(p_{j\alpha} r_{j\epsilon} + p_{j\epsilon} r_{j\alpha}) \delta_{\beta\gamma} + (p_{j\beta} r_{j\epsilon} + p_{j\epsilon} r_{j\beta}) \delta_{\gamma\alpha} \\ & + (p_{j\gamma} r_{j\epsilon} + p_{j\epsilon} r_{j\gamma}) \delta_{\alpha\beta} + (p_{j\beta} r_{j\gamma} + p_{j\gamma} r_{j\beta}) \delta_{\alpha\epsilon} \\ & + (p_{j\alpha} r_{j\gamma} + p_{j\gamma} r_{j\alpha}) \delta_{\beta\epsilon} + (p_{j\alpha} r_{j\beta} + p_{j\beta} r_{j\alpha}) \delta_{\gamma\epsilon}] \\ & - 10 (\mathbf{r}_j \cdot \mathbf{p}_j) (r_{j\alpha} r_{j\epsilon} \delta_{\beta\gamma} + r_{j\beta} r_{j\epsilon} \delta_{\gamma\alpha} + r_{j\gamma} r_{j\epsilon} \delta_{\alpha\beta} \\ & + r_{j\beta} r_{j\gamma} \delta_{\alpha\epsilon} + r_{j\alpha} r_{j\gamma} \delta_{\beta\epsilon} + r_{j\alpha} r_{j\beta} \delta_{\gamma\epsilon}) \\ & + 4 (\mathbf{r}_j \cdot \mathbf{p}_j) {r_j}^2 (\delta_{\alpha\epsilon} \delta_{\beta\gamma} + \delta_{\beta\epsilon} \delta_{\gamma\alpha} + \delta_{\gamma\epsilon} \delta_{\alpha\beta}) \}. \end{aligned} \tag{A6}$$

The SAMM$_p$ algorithm treats these PMM multipole moments $\tilde{\mathbf{M}}^{m,\mathbf{0}}$ exactly like in the MM case, i.e., they are shifted to a different reference point using Eq. (19) of Ref. 47, and potentials and Taylor expansion coefficients are calculated from Eqs. (9) and (10) of Ref. 47, respectively. In IPHIGENIE, the electrostatics of static partial charges is calculated only once for each integration step and is reused during the PMM-SCF iteration.

[1] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[2] H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[3] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[4] A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguven, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[5] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).
[6] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, J. Comput. Chem. **25**, 1656 (2004).
[7] P. Tavan, H. Carstens, and G. Mathias, "Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives," in *Protein Folding Handbook*, edited by J. Buchner and T. Kiefhaber (Wiley-VCH, Weinheim, 2005), Vol. 1, pp. 1170–1195.
[8] M. A. Thompson and G. K. Schenter, J. Phys. Chem. **99**, 6374 (1995).
[9] M. A. Thompson, J. Phys. Chem. **100**, 14492 (1996).
[10] D. Bakowies and W. Thiel, J. Phys. Chem. **100**, 10580 (1996).
[11] J. Gao, J. Comput. Chem. **18**, 1061 (1997).
[12] D. P. Geerke, S. Thiel, W. Thiel, and W. F. van Gunsteren, J. Chem. Theory Comput. **3**, 1499 (2007).
[13] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[14] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[15] U. C. Singh and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986).
[16] G. Jansen, F. Colonna, and J. G. Ángyán, Int. J. Quantum Chem. **58**, 251 (1996).
[17] M. J. Field, Mol. Phys. **91**, 835 (1997).
[18] I. H. Hillier, J. Mol. Struct.: THEOCHEM **463**, 45 (1999).
[19] H. Houjou, Y. Inoue, and M. Sakurai, J. Phys. Chem. B **105**, 867 (2001).
[20] L. Jensen, P. T. van Duijnen, and J. G. Snijders, J. Chem. Phys. **118**, 514 (2003).
[21] C. J. R. Illingworth, S. R. Gooding, P. J. Winn, G. A. Jones, G. G. Ferenczy, and C. A. Reynolds, J. Phys. Chem. A **110**, 6487 (2006).
[22] M. Wanko, M. Hoffmann, J. Fraehmcke, T. Frauenheim, and M. Elstner, J. Phys. Chem. B **112**, 11468 (2008).
[23] C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes, and B. Mennucci, J. Chem. Theory Comput. **5**, 1838 (2009).
[24] R. A. Bryce, M. A. Vincent, N. O. J. Malcolm, I. H. Hillier, and N. A. Burton, J. Chem. Phys. **109**, 3077 (1998).
[25] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, J. Phys. Chem. A **105**, 293 (2001).
[26] M. Dupuis, M. Aida, Y. Kawashima, and K. Hirao, J. Chem. Phys. **117**, 1242 (2002).
[27] H. Li and M. S. Gordon, J. Chem. Phys. **126**, 124112 (2007).
[28] P. K. Biswas and V. Gogonea, J. Chem. Phys. **129**, 154108 (2008).
[29] F. Lipparini, C. Cappelli, G. Scalmani, N. De Mitri, and V. Barone, J. Chem. Theory Comput. **8**, 4270 (2012).
[30] Z. Lu and Y. Zhang, J. Chem. Theory Comput. **4**, 1237 (2008).
[31] E. Boulanger and W. Thiel, J. Chem. Theory Comput. **8**, 4527 (2012).
[32] K. A. Maerzke, G. Murdachaew, C. J. Mundy, G. K. Schenter, and J. I. Siepmann, J. Phys. Chem. A **113**, 2075 (2009).
[33] G. Murdachaew, C. J. Mundy, and G. K. Schenter, J. Chem. Phys. **132**, 164102 (2010).
[34] M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello, J. Chem. Phys. **110**, 10452 (1999).
[35] M. Nonella, G. Mathias, and P. Tavan, J. Phys. Chem. A **107**, 8638 (2003).
[36] M. Schmitz and P. Tavan, J. Chem. Phys. **121**, 12247 (2004).
[37] V. Schultheis, R. Reichold, B. Schropp, and P. Tavan, J. Phys. Chem. B **112**, 12217 (2008).
[38] A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
[39] T. Laino, F. Mohamed, A. Laio, and M. Parrinello, J. Chem. Theory Comput. **1**, 1176 (2005).
[40] G. Babitzki, R. Denschlag, and P. Tavan, J. Phys. Chem. B **113**, 10483 (2009).
[41] G. Babitzki, G. Mathias, and P. Tavan, J. Phys. Chem. B **113**, 10496 (2009).
[42] B. Rieff, S. Bauer, G. Mathias, and P. Tavan, J. Phys. Chem. B **115**, 11239 (2011).
[43] M. Nonella, G. Mathias, M. Eichinger, and P. Tavan, J. Phys. Chem. B **107**, 316 (2003).
[44] M. Klähn, G. Mathias, C. Kotting, J. Schlitter, M. Nonella, K. Gerwert, and P. Tavan, J. Phys. Chem. A **108**, 6186 (2004).
[45] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[46] J. VandeVondele, P. Tröster, P. Tavan, and G. Mathias, J. Phys. Chem. A **116**, 2466 (2012).
[47] K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, and P. Tavan, J. Chem. Theory Comput. **8**, 3628 (2012).

[48]J. Hutter, A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman, and M. Parrinello, CPMD: Car–Parinello Molecular Dynamics, Version 3.15.1, © IBM Corp, 1990–2008 and MPI für Festkörperforschung Stuttgart, 1997–2001, see www.cpmd.org.
[49]J. Applequist, J. R. Carl, and K.-K. Fung, J. Am. Chem. Soc. **94**, 2952 (1972).
[50]P. Ahlström, A. Wallquist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989).
[51]P. E. Lopes, B. Roux, and A. D. MacKerell, Jr., Theor. Chem. Acc. **124**, 11 (2009).
[52]B. T. Thole, Chem. Phys. **59**, 341 (1981).
[53]D. Elking, T. Darden, and R. J. Woods, J. Comput. Chem. **28**, 1261 (2007).
[54]S. Russell and A. Warshel, J. Mol. Biol. **185**, 389 (1985).
[55]D. V. Belle, I. Couplet, M. Prevost, and S. J. Wodak, J. Mol. Biol. **198**, 721 (1987).
[56]P. K. Biswas and V. Gogonea, J. Chem. Phys. **123**, 164114 (2005).
[57]M. P. Allen and D. Tildesley, in *Computer Simulations of Liquids* (Clarendon, Oxford, 1987), Chap. 1.5.2, pp. 24–27.
[58]C. Niedermeier and P. Tavan, J. Chem. Phys. **101**, 734 (1994).
[59]C. Niedermeier and P. Tavan, Mol. Simul. **17**, 57 (1996).
[60]G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[61]A. A. Appel, SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **6**, 85 (1985).
[62]J. Barnes and P. Hut, Nature (London) **324**, 446 (1986).
[63]L. Greengard and V. Rokhlin, J. Comput. Phys. **73**, 325 (1987).
[64]W. Dehnen, J. Comput. Phys. **179**, 27 (2002).
[65]M. Eichinger, H. Grubmüller, H. Heller, and P. Tavan, J. Comput. Chem. **18**, 1729 (1997).
[66]D. Carfi, AAPP Phys. Math. Nat. Sci. **88**, C1A1001004 (2010).
[67]A. D. Buckingham, Adv. Chem. Phys. **12**, 107 (1967).
[68]K. Hinsen and B. U. Felderhof, J. Math. Phys. **33**, 3731 (1992).
[69]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1988), Chap. 3.2, pp. 118–120.
[70]P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[71]J. Hutter, H. P. Lüthi, and M. Parrinello, Comput. Mater. Sci. **2**, 244 (1994).
[72]L. Verlet, Phys. Rev. **159**, 98 (1967).
[73]V. Kräutler, W. F. van Gunsteren, and P. Hünenberger, J. Comput. Chem. **22**, 501 (2001).
[74]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[75]J. Perdew and W. Yue, Phys. Rev. B **33**, 8800 (1986).
[76]N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[77]W. S. Benedict, N. Gailar, and E. K. Plyler, J. Chem. Phys. **24**, 1139 (1956).
[78]K. Ichikawa, Y. Kameda, T. Yamaguchi, H. Wakita, and M. Misawa, Mol. Phys. **73**, 79 (1991).
[79]W. F. Murphy, J. Chem. Phys. **67**, 5877 (1977).
[80]T. Dyke and J. Muenter, J. Chem. Phys. **59**, 3125 (1973).
[81]R. A. Buckingham and J. Corner, Proc. R. Soc. London, Ser. A **189**, 118 (1947).
[82]G. Kell, J. Chem. Eng. Data **12**, 66 (1967).
[83]G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[84]J. L. F. Abascal and C. Vega, J. Chem. Phys. **123**, 234505 (2005).
[85]M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, 1st ed. (Oxford University Press, New York, 2010), Chap. 3.13, pp. 121–124.
[86]See supplementary material at http://dx.doi.org/10.1063/1.4811292 for a total of four figures (S9– S12) and three equations (S39– S41) as additional material to the results of the (V) test simulations, on five pages in three sections.
[87]J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).
[88]G. Mathias and M. D. Baer, J. Chem. Theory Comput. **7**, 2028 (2011).
[89]J. Hutter and A. Curioni, Parallel Comput. **31**, 1 (2005).
[90]P. Tröster, K. Lorenzen, M. Schwörer, and P. Tavan, "Polarizable water models from mixed computational and empirical optimization" J. Phys. Chem. B (to be published).

Der folgende Abdruck

Supplementary Material to
Coupling DFT to polarizable force fields for efficient and accurate
Hamiltonian molecular dynamics simulations

Magnus Schwörer, Benedikt Breitenfeld, Philipp Tröster, Sebastian Bauer,
Konstantin Lorenzen, Paul Tavan und Gerald Mathias

J. Chem. Phys. **138**, 244103 (2013)

liefert zusätzliche Informationen zum vorangestellten Haupttext. Abschnitt S1 dokumentiert die Energieerhaltung der PMM-Implementierung in IPHIGENIE, Abschnitt S2 zeigt dass die Energiefluktuationen in DFT/PMM-MD vergleichbar mit denen in DFT-MD sind, und in Abschnitt S3 wird der Einfluß der Wahl des DFT-Selbstkonsistenzkriteriums auf das Dipolmoment des DFT-Fragments untersucht.

Supplementary material to:

Coupling DFT to polarizable force fields for efficient and accurate Hamiltonian molecular dynamics simulations

Magnus Schwörer, Benedikt Breitenfeld, Philipp Tröster, Sebastian Bauer, Konstantin Lorenzen, Paul Tavan, and Gerald Mathias[a)]

Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

(Dated: 5 April 2013)

[a)]Electronic mail: gerald.mathias@physik.uni-muenchen.de

S1. PMM ENERGY CONSERVATION

A numerical integration of a Hamiltonian dynamics with the Verlet algorithm[1] does not exactly conserve the energy. Instead, for any given size of the time step Δt the energy is given by[2]

$$E(t\,|\,\Delta t) = \langle E(t\,|\,\Delta t)\rangle_\tau + f(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{S39}$$

where $\langle E(t\,|\,\Delta t)\rangle_\tau$ is the temporal average of the energy over a reasonably large time span τ, which converges in the analytical limit $\Delta t \to 0$ to the exact energy E_0, and where the corresponding average of the shape function $f(t)$ vanishes, i.e. $\lim_{\tau\to\infty}\langle f(t)\rangle_\tau = 0$. Then the energy fluctuations $\Delta E(t\,|\,\Delta t) \equiv E(t\,|\,\Delta t) - \langle E(t\,|\,\Delta t)\rangle_\tau$ are given by

$$\Delta E(t\,|\,\Delta t) = f(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{S40}$$

with the standard deviation

$$\sigma_{\Delta E} = \Delta t^2 \sqrt{\langle f^2(t)\rangle_\tau} + \mathcal{O}(\Delta t^4), \tag{S41}$$

which vanishes in the analytical limit $\Delta t \to 0$.

Equations (S40) and (S41) should apply to the numerically integrated Hamiltonian dynamics of the PMM water dimer. For a check whether our implementation is correct, we consider two TL4P$_{\text{ini}}$ water models (Tröster et al., in preparation). These stiff models are

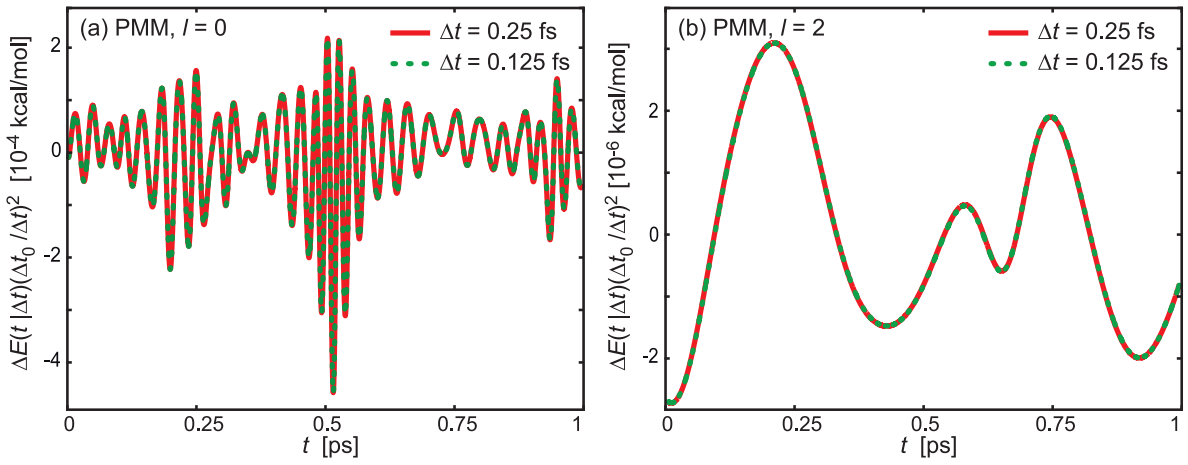


FIG. S9. Fluctuations of the energy of a PMM water dimer for different time steps $\Delta t = 0.25$ fs (red) and $\Delta t = 0.125$ fs (dotted green); (a) dimer at close contact ($l = 0$), (b) dimer softy restrained to $d_{\text{OO}} \approx 10$ Å ($l = 2$).

either positioned at close contact, such that the electrostatics is calculated from the usual Coulomb expressions (and the extensions to Gaussian dipoles) and the van der Waals interactions from Buckingham potentials[3] centered at the oxygen atoms, or are softly restrained to distances of about 10 Å, at which the electrostatics is treated by the $SAMM_4$ algorithm[4] and the van der Waals interaction is neglected.

If one integrates the Hamiltonian dynamics by the Verlet algorithm for the two dimer-setups starting at identical initial conditions but using different time steps (e.g. $\Delta t = \Delta t_0$ and $\Delta t = \Delta t_0/2$), then the scaled energy fluctuations $\Delta E(t\,|\,\Delta t)(\Delta t_0/\Delta t)^2$ should have according to Eq. (S40) identical shapes $f(t)\Delta t_0^2$. Figure S9 shows the scaled energy fluctuations $\Delta E(t\,|\,\Delta t)(\Delta t_0/\Delta t)^2$ observed in the dimer simulations for a basic time step $\Delta t_0 = 0.25$ fs. The fluctuations $\Delta E(t\,|\,\Delta t)$ observed for $\Delta t = \Delta t_0$ are drawn by a red solid line, the scaled observations $4\Delta E(t\,|\,\Delta t)$ observed for $\Delta t = \Delta t_0/2$ by a green dotted line. After scaling the fluctuations are seen to match almost perfectly for the two dimer setups characterized further above as documented by Figs. S9(a) and S9(b).

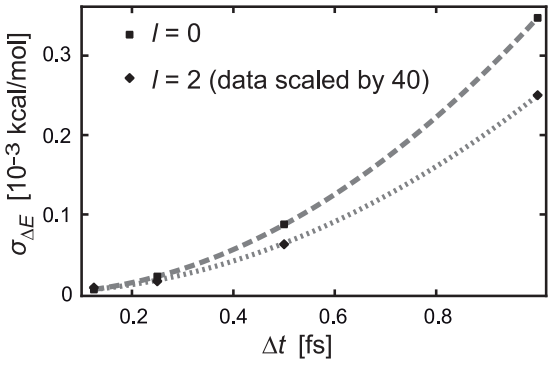


FIG. S10. Standard deviations $\sigma_{\Delta E}$ of the energy fluctuations $\Delta E(t\,|\,\Delta t)$ for different time steps $\Delta t \in \{0.125, 0.250, 0.500, 1.000\}$. Results for the dimer settings $l = 0$ and $l = 2$ are depicted by squares and diamonds, respectively. The gray dashed lines show fitted parabolas. For $l = 2$ scaled values $40\sigma_{\Delta E}$ are drawn for purposes of improved visibility.

According to Eq. (S41) the standard deviation $\sigma_{\Delta E}$ of the energy fluctuations $\Delta E(t\,|\,\Delta t)$ should vanish quadratically with the time step size Δt. Figure S10 shows that this is actually the case as demonstrated by MD simulations with the time steps $\Delta t \in \{0.125, 0.25, 0.50, 1.00\}$ of the PMM dimer at close contact ($l = 0$) and softly restrained to an average distance $d_{\mathrm{OO}} = 10$ Å ($l = 2$).

S2. DFT/PMM ENERGY CONSERVATION FOR THE DIMER

The main text compares for the dimer at close contact the energy fluctuations of the DFT reference simulation in Fig. 4(c) with the fluctuations of the corresponding DFT/PMM simulations in Fig. 5(a). A part of these data is repeated at a much higher time resolution (80 fs instead of 2 ps) in Figures S11(a) and S11(b). The comparison at the enhanced time resolution demonstrates that the DFT and DFT/PMM fluctuations have similar amplitudes and frequencies. In addition the Figures S11(c) and S11(d) show part of the data presented

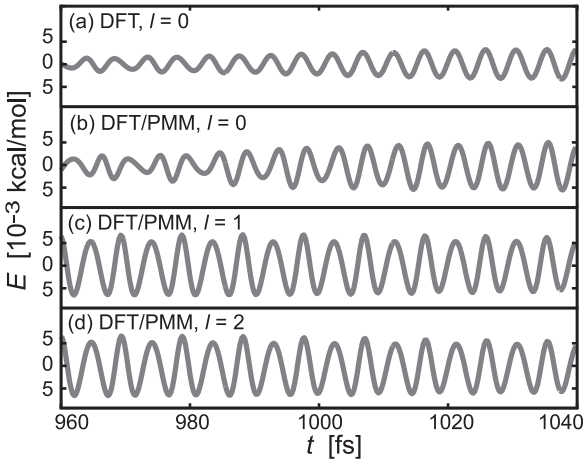


FIG. S11. Temporally resolved energy conservation in the DFT reference and DFT/PMM hybrid dimer simulations. (a) Close contact DFT, (b) Close contact DFT/PMM ($l = 0$), (c) DFT/PMM softly restrained to $d_{\mathrm{OO}} \approx 7$ Å ($l = 1$), and (d) to $d_{\mathrm{OO}} \approx 10$ Å ($l = 2$).

at a lower time resolution in Figs. 5(b) and 5(c), respectively, at the enhanced time resolution. The depicted fluctuations refer to the DFT/PMM dimer softly restrained to average distances $d_{\mathrm{OO}} \approx 7$ Å and $d_{\mathrm{OO}} \approx 10$ Å, respectively. As compared to close contact they have slightly larger amplitudes but similar frequencies.

S3. DFT CONVERGENCE CRITERION FOR DFT/PMM-MD

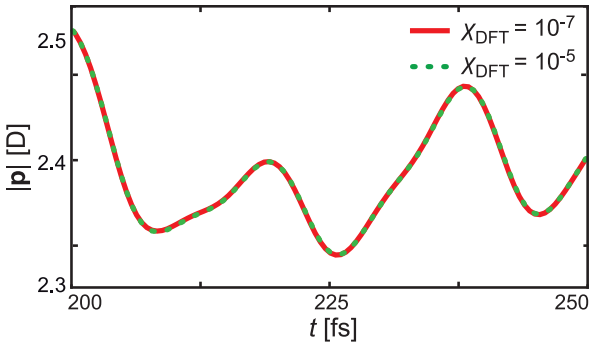


FIG. S12. Invariance of two 250 fs DFT/PMM-MD liquid water trajectories (started at identical initial conditions) with respect to increasing χ_{DFT} from 10^{-7} (red curve) to 10^{-5} (green dotted curve). The sample observable is the absolute value $|\mathbf{p}(t)|$ of the DFT fragment's dipole moment.

Figure S12 illustrates for the absolute value of the dipole moment $|\mathbf{p}(t)|$ of a DFT water molecule embedded in a PMM water environment that the DFT convergence criterion χ_{DFT} can be safely loosened from 10^{-7} to 10^{-5} without changing DFT/PMM trajectories. Similar identities result for the components of $\mathbf{p}(t)$ and for the DFT fragment's energy $E(t)$. The identities of these trajectories (and the reduced computational effort) motivate the choice $\chi_{\mathrm{DFT}} = 10^{-5}$ as default for DFT/(P)MM-MD simulations. Note that $\chi_{\mathrm{DFT}} = 10^{-5}$ is also the default in CPMD.[5]

REFERENCES

[1]L. Verlet, Phys. Rev. **165**, 201 (1967).

[2]M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, 1st ed. (Oxford University Press USA, New York, 2010) Chap. 3.13, pp. 121–124.

[3]R. A. Buckingham and J. Corner, Proc. R. Soc. London, Ser. A **189**, 118 (1947).

[4]K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, and P. Tavan, J. Chem. Theory Comput. **8**, 3628 (2012).

[5]J. Hutter, A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman, and M. Parrinello, *CPMD: Car–Parinello Molecular Dynamics, version 3.15.1*, © IBM Corp 1990–2008 and MPI für Festkörperforschung Stuttgart 1997–2001, `www.cpmd.org`.

2.2 Steigerung der Effizienz von DFT/PMM

Die nachfolgend abgedruckte Publikation

Utilizing Fast Multipole Expansions for Efficient and Accurate
Quantum-Classical Molecular Dynamics Simulations

Magnus Schwörer, Konstantin Lorenzen, Gerald Mathias und Paul Tavan
J. Chem. Phys. **142**, 104108 (2015),

die ich zusammen mit Konstantin Lorenzen, Gerald Mathias und Paul Tavan verfasst habe, zeigt, wie sich die Performanz der DFT/PMM-Kopplung durch Anwendung des mittlerweile entwickelten neuen SAMM$^{\chi}$-Algorithmus [109, 127] um eine Größenordnung erhöhen lässt. Dazu wird einer neuen Ebene von Subvolumina des DFT-Fragments (Voxel) eingeführt, und die tatsächliche Ausdehnung der DFT-Ladungsverteilung bei der Berechnung der für SAMM nötigen Gyrationsradien berücksichtigt. Ferner wird die Integration des DFT-Programms CPMD [152] in das PMM-MD-Programm IPHIGENIE beschrieben und das Skalierungsverhalten des neuen Hybridprogramms IPHIGENIE/CPMD auf Großrechnern untersucht.

Utilizing fast multipole expansions for efficient and accurate quantum-classical molecular dynamics simulations

Magnus Schwörer, Konstantin Lorenzen, Gerald Mathias, and Paul Tavan[a)]
Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

(Received 16 January 2015; accepted 25 February 2015; published online 12 March 2015)

Recently, a novel approach to hybrid quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) simulations has been suggested [Schwörer *et al.*, J. Chem. Phys. **138**, 244103 (2013)]. Here, the forces acting on the atoms are calculated by grid-based density functional theory (DFT) for a solute molecule and by a polarizable molecular mechanics (PMM) force field for a large solvent environment composed of several 10^3-10^5 molecules as negative gradients of a DFT/PMM hybrid Hamiltonian. The electrostatic interactions are efficiently described by a hierarchical fast multipole method (FMM). Adopting recent progress of this FMM technique [Lorenzen *et al.*, J. Chem. Theory Comput. **10**, 3244 (2014)], which particularly entails a strictly linear scaling of the computational effort with the system size, and adapting this revised FMM approach to the computation of the interactions between the DFT and PMM fragments of a simulation system, here, we show how one can further enhance the efficiency and accuracy of such DFT/PMM-MD simulations. The resulting gain of total performance, as measured for alanine dipeptide (DFT) embedded in water (PMM) by the product of the gains in efficiency and accuracy, amounts to about one order of magnitude. We also demonstrate that the jointly parallelized implementation of the DFT and PMM-MD parts of the computation enables the efficient use of high-performance computing systems. The associated software is available online. [http://dx.doi.org/10.1063/1.4914329]

I. INTRODUCTION

We have recently presented a novel quantum-classical[1–3] approach for molecular dynamics (MD) simulations, which combines density functional theory[4,5] (DFT) for a subsystem with (polarizable) molecular mechanics [(P)MM] force fields atomistically modeling its condensed phase environment.[6] Like other schemes, which combine high-level quantum mechanical (QM) methods with PMM force fields for hybrid MD simulations,[7–13] our DFT/PMM approach also provides the required analytical gradients. Our hybrid method, in particular, accomplishes for sizeable DFT fragments an efficient treatment of large condensed-phase PMM environments by applying a fast multipole method[14–21] (FMM) to the electrostatic interactions.[6,22]

More specifically, our DFT/PMM scheme combines the grid-based and parallelized plane-wave DFT program CPMD[23] with the parallelized PMM-MD program IPHIGENIE[6,21,24–26] and enables efficient and accurate Hamiltonian MD simulations of such hybrid DFT/PMM systems. Applying toroidal boundary conditions,[27] the long-range electrostatics is treated by the fourth order structure-adapted fast multipole method,[21] SAMM$_4$, within interaction spheres of radius d_{MIC} surrounding the atoms. For distances beyond d_{MIC}, which is given by the minimum image convention[27] and is about half the size of the periodic unit cell, the electrostatics is described by a reaction field (RF) approach.[20]

As discussed in detail by Lorenzen *et al.*,[24] the resulting SAMM$_4$/RF algorithm is a non-periodic alternative to the lattice summations (LS) of the Ewald type[28–30] for the treatment of the long-range electrostatics in MM[31] and QM/MM[32–35] MD simulations. Replacing the RF model, SAMM$_4$ could alternatively be combined with a LS approach, treating the interactions among the Cartesian point multipoles[36] of the top-level clusters and all their periodic images with a real-space cutoff as large as d_{MIC}. Because both alternatives exclusively deal with the multipole moments of the few top-level clusters, they can be equally well applied to (P)MM and DFT/(P)MM systems.

Like its predecessors,[17–19] SAMM$_4$/RF also maps the atoms of a simulation system onto a quaternary tree of nested atomic clusters, whose lowest level consists of fixed chemical motifs, the so-called structural units (SUs), whereas its higher levels are adaptively formed by self-organizing clustering algorithms.[24] Furthermore, also SAMM$_4$ applies fixed distance classes to the decisions, whether large clusters of atoms should be decomposed into the enclosed sub-clusters for the hierarchical computation of the electrostatics, which characterizes FMM.[14–16] Here, particularly the innermost class boundary, which is marked by SU distances smaller than a limiting distance d_0, signifies the transition from the approximate FMM treatment of the electrostatic interactions among SUs to exact descriptions for their constituent atoms. In SAMM$_4$, this distance d_0 concurrently represents the cutoff distance for the van der Waals interactions, for which values in the range of about 10-15 Å are common choices in force

a)Electronic mail: paul.tavan@physik.uni-muenchen.de

fields[37–41] for bio-molecular simulation. Correspondingly, SAMM$_4$ also employs values for d_0 from this range.[21]

However, the use of fixed distance classes does not take full advantage of the FMM concept, whose expansion parameter is the apparent size $\vartheta \equiv 2\tilde{R}/r$ at which a cluster of atoms of diameter $2\tilde{R}$ appears at a distance r. If one wants to concurrently optimize the computational accuracy and efficiency, i.e., the performance, then the so-called interaction acceptance criterion (IAC), upon which the algorithm decides whether a pair of clusters is small and distant enough for a sufficiently accurate FMM description, should be based on their apparent sizes ϑ.

By extending the FMM scheme toward q'th order expansions ($q = 1,2,3$) of the dispersion attraction ($\sim -1/r^6$) and subsequently, also to first order expansions for the Lennard-Jones model ($\sim 1/r^{12}$) of the Pauli repulsion, SAMM$_4$ has been recently revised toward the use of an IAC, which now actually employs the apparent size ϑ for decisions.[24,42] Additionally, the SAMM algorithms were augmented by representing the FMM forces as exact derivatives of the FMM energy expressions.[42] A corresponding algorithm is marked by an additional superscript "E". Three variants $\chi \in \{\mathrm{f,m,a}\}$ of the thus obtained SAMM$^{\chi,\mathrm{E}}_{4,q,1}$/RF algorithm were defined, where the label χ characterizes the performances ranging from "fast and still reasonably accurate" (χ = f) over "intermediate" (χ = m) to "very accurate but comparatively slow" (χ = a). Because the apparent size ϑ is calculated from accuracy weighted cluster radii $\tilde{R} \equiv R/a$, where R is a cluster's radius of gyration and $a \geq 1$ an accuracy correction factor, these performances are almost independent of the chemical composition of the respective simulation system.[24]

As demonstrated for MM and PMM sample simulation systems,[24,42] the accuracy and efficiency of these new SAMM$^{\chi,\mathrm{E}}_{4,q,1}$/RF algorithms for the computation of the electrostatic and van der Waals interactions clearly outperforms the preceding distance class algorithm SAMM$_4$/RF. Furthermore, a strict linear scaling of the computational effort has been demonstrated[24] for N-atomic systems with N up to 10^5. In the case of liquid water, the treatment of the complex PMM model potential, which features five point charges and an inducible Gaussian dipole and is called TL6P,[43] turned out to be only by a factor five more expensive than the simulation of the simple three-point MM model, TIP3P.[44]

The strongly improved performance of the revised and extended SAMM$^{\chi,\mathrm{E}}_{4,q,1}$ algorithms immediately suggests that their integration into the hybrid DFT/PMM scheme[6] mentioned above should also lead to a likewise enhanced performance here. It is the purpose of this contribution to show that this aim can be actually reached, if the previous computational strategy of the DFT/PMM interface is thoroughly revised toward an IAC, which is based on apparent sizes ϑ.

For distant atomic clusters, this revision will turn out to be almost trivial, if one assigns suitable sizes to clusters containing DFT atoms. Then, one simply can replace the SAMM$_4$ distance class scheme[21] by the novel IAC of SAMM$^{\chi,\mathrm{E}}_{4,q,1}$, which considers apparent sizes ϑ.[24] For the lower interaction levels, however, which treat the electrostatic interactions of individual PMM atoms with the charge density distributed on the DFT grid,[6] the new IAC offers the possibility to extend the FMM concepts toward hierarchically nested sub-volumes of the DFT grid.

Following this route, one should be able to achieve large performance gains, because it entails the chance to strongly reduce the number of PMM atoms, for which one has to compute the electrostatic interactions with the huge number $\Gamma > 10^6$ of DFT grid charges q_γ ($\gamma = 1,\ldots,\Gamma$) by means of costly exact pair expressions. Combining atomic sized sub-volumes of the DFT grid with an IAC based on their apparent size will leave only very few PMM atoms at distances too small for the much more efficient calculation of the interactions by FMM expressions.

In our following presentation of the revised interface between the DFT program CPMD[23] and the PMM-MD program IPHIGENIE,[6,21,24,25] we first rehash the basic concepts leading to a symmetric and essentially energy conserving FMM treatment of the electrostatics in a DFT/PMM setting.[6] Subsequently, we sketch the new hierarchical partitioning of the DFT grid, which favorably exploits the new IAC offered by SAMM$^{\chi,\mathrm{E}}_{4,q,1}$.[24] Next, we introduce as a sample simulation system, a DFT dipeptide (Ac-Ala-NHMe) embedded in liquid PMM water as modeled by the TL6P potential[43] and describe the observables employed for checks of efficiency and accuracy. For this sample system, we study issues of algorithmic performance in MD simulations. After a few remarks on the jointly parallel implementation of IPHIGENIE and CPMD and on its scaling properties on high-performance computing (HPC) systems, we summarize the achievements resulting from the revision of the interface.

II. THEORY

An optimally performing and Hamiltonian computation of the electrostatic interactions between the DFT and PMM fragments is a key task in corresponding hybrid methods.[6] We start with a short introduction into the algorithmic concepts employed in FMM combinations of grid-based DFT with PMM force fields and subsequently turn to the revisions aiming at an enhanced performance.

A. Basics of DFT/PMM electrostatic interactions

The electrostatic DFT/PMM interaction energy $H^{\mathrm{DFT/PMM}}_{\mathrm{elec}}$ can be decomposed by

$$H^{\mathrm{DFT/PMM}}_{\mathrm{elec}} = \int d\mathbf{r}\,[\rho_{\mathrm{e}}(\mathbf{r}) + \rho_{\mathrm{c}}(\mathbf{r})]\,\Phi(\mathbf{r}) \qquad (1)$$

into the energies of the DFT fragment's electron (ρ_{e}) and core (ρ_{c}) charge densities in the external potential Φ generated by the PMM fragment. Here, solely the computation of the electronic contribution poses a challenge.[6]

In grid-based programs like CPMD,[23] $\rho_{\mathrm{e}}(\mathbf{r})$ is represented by point charges q_γ residing at the locations $\mathbf{r}_\gamma$ of all Γ grid points γ. For a given DFT box, which we initially assume to be cubic, the spatial resolution of the grid, and thus, the number Γ are determined by the plane-wave cutoff energy E_{cut}. Typically, one has $\Gamma \approx 10^6$. The electronic contribution to Eq. (1) is then the extended sum $\sum_\gamma q_\gamma \Phi(\mathbf{r}_\gamma)$.

Thus, in a DFT/PMM computation, the external potential Φ has to be evaluated at all points $\mathbf{r}_\gamma$. To avoid singularities of Φ within the volume V of the DFT box, which would arise from enclosed PMM point charges or dipoles, smooth models[6,22,45–47] have to be used for generating static partial charges q_i and induced dipoles $\mathbf{p}_i$ of the PMM atoms i. For point-like objects at $\mathbf{r}_i$, a possible choice is the smoothing by a Gaussian shape function[6,22,46]

$$g(\mathbf{r} \mid \mathbf{r}_i, \sigma_i) = \frac{1}{(2\pi\sigma_i^2)^{3/2}} \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{2\sigma_i^2}\right]. \tag{2}$$

Here, different widths $\sigma_i \lesssim 0.8$ Å may be employed for the various PMM charges and induced dipoles.[48] The Gaussian character can be approximately neglected as soon as the scaled distance $|\mathbf{r}-\mathbf{r}_i|/\sigma_i$ exceeds a certain threshold n, which we will determine by an analysis of the Gaussian truncation errors.

Once all values $\Phi(\mathbf{r}_\gamma)$ have been calculated for a given static configuration of the system, the iterative computation of the Kohn-Sham (KS) orbitals[5] can be initiated. In a joint iteration, which involves the repeated computation of the mutual polarization of the grid charges and of the induced PMM dipoles, the KS orbitals and the induced dipoles have to be brought to self-consistency.[6] Finally, the converged DFT charge distribution yields the forces on the M DFT atoms μ ($\mu = 1, \ldots, M$) and the electrostatic reaction forces on the PMM atoms i as the negative gradients of the Hamiltonian (1) with respect to the atomic positions.

Without further provisions the above tasks would involve as many as 10^{10} distance calculations already for small DFT and PMM fragments. Integrating the DFT fragment into a FMM scheme can substantially reduce this computational effort by more than two orders of magnitude.[6,22,45] Here, we will now show that another large performance gain can be achieved, if the DFT/PMM electrostatics is systematically adapted to the most recent FMM approach provided by $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$.[24,42]

B. $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$

$\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$ collects atoms into hierarchically nested clusters of increasing size and calculates their electrostatic interactions by balanced m'th order multipole and n'th order local Taylor expansions up to order $m+n=4$ around their centers of geometry. Similarly, the dispersion attraction is treated by FMM expansions up to order $q \le 3$,[24] whereas the expansions of the Pauli repulsion are restricted to order 1.[42] The resulting forces obey Newton's reaction principle[21] and conserve the FMM energy.[42] The removed 10 Å cutoff of the van der Waals interactions leads to a strongly enhanced accuracy and enables a decrease of the minimum distance d_0, at which two clusters of a PMM simulation system are allowed to interact at the lowest level of the FMM cluster hierarchy, to about 5.5 - 7.5 Å. Because the computational effort is dominated by the costly atomic pair interactions, which have to be evaluated for distances smaller than d_0, the reduced d_0 leads for $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ to much smaller numbers of such interactions and, thus, to speedup factors in the range from 5 (χ = f) to 2 (χ = a).

The quality of the $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$ approximations for the interaction of two clusters A and B, whose geometric centers are a distance r apart, is controlled by the IAC formula[24]

$$\frac{1}{2}\left[\vartheta_{A,l}(r) + \vartheta_{B,l}(r)\right] \le \Theta_\chi, \tag{3}$$

where Θ_χ is a certain threshold and

$$\vartheta_{A,l}(r) \equiv \frac{1}{a_{A,l}} \frac{2R_{A,l}}{r} \tag{4}$$

is the accuracy weighted apparent size of a cluster A on FMM cluster level l, whose radius of gyration is denoted by $R_{A,l}$. The accuracy correction $1 \le a_{A,l} \le 1.8$ depends, for (P)MM clusters, on their geometry and electrostatic signature.[24] For DFT clusters A, which occupy a SAMM cluster level $l \le 0$, we choose $a_{A,l} = 1$. Note that here we employ the SAMM numbering[19,21] of clustering levels according to which the level $l = 0$ is composed of predefined SUs comprising 3 to 16 atoms (for most recent definitions, see Ref. 24). Levels $l > 0$ contain compact clusters comprising on average four clusters collected from level $l-1$, whereas the level $l = -1$ consists of individual atoms.[24]

Having chosen a certain threshold Θ_χ, $\chi \in \{\mathrm{f}, \mathrm{m}, \mathrm{a}\}$, the interaction between two clusters A and B is calculated on hierarchy level l, if the IAC (3) holds. Otherwise, the two clusters are decomposed into their children and (3) is checked again for the children interactions on level $l-1$. $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$ offers the accuracy thresholds $\Theta_\mathrm{f} = 0.25$, $\Theta_\mathrm{m} = 0.20$, and $\Theta_\mathrm{a} = 0.17$, which represent different and quantitatively well-defined compromises between efficiency and accuracy.[24]

C. Exploiting $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$ for DFT/PMM interactions

According to the IAC (3) and the definition (4) of the apparent size $\vartheta_{A,l}(r)$ of a cluster A, the interaction of two clusters on level l can be efficiently calculated by FMM down to very small cluster-cluster distances r, if the radii of gyration are correspondingly small. Consider now the electrostatic interaction of a PMM atom i with a nearby DFT atom μ and assume that the enclosing SUs fail to pass the IAC (3) at level $l = 0$. In a pure PMM setting, this interaction would then have to be treated at the SAMM level $l = -1$ using exact pair expressions.

In a DFT/PMM setting, however, we have to compute the electrostatic potential $\Phi(\mathbf{r})$ generated by the PMM atoms i not only at the positions $\mathbf{r}_\mu$ of the few DFT atoms μ, but additionally at all points $\mathbf{r}_\gamma$ of the whole DFT grid. Given a predefined accuracy level χ, this task is executed at maximal computational efficiency, if FMM expansions are used for as many grid points γ as possible.

For an IAC based on apparent sizes ϑ, the elementary FMM interaction partners of the PMM atoms i should thus be small and equally sized portions of the spatially quite extended distribution of grid charges q_γ. We define these portions by commensurably partitioning the DFT grid into sub-volumes V_λ ($\lambda = 1, \ldots, \Lambda$) of atomic size, which we call voxels. For a cubic DFT grid, the voxels are then also cubic. Measuring the common size of the voxels by their radius of gyration R_Λ and noting that the radius of gyration R_i of a PMM atom i vanishes,

Eqs. (4) and (3) demonstrate that i does not contribute to the IAC.

For this case, the general FMM IAC (3) should be modified to the so-called FMM voxel IAC,

$$2R_\Lambda/r_{\lambda i} \le \Theta_\chi, \tag{5}$$

where we have set the accuracy weight in (4) to one and have replaced the average radius of gyration by R_Λ. Because R_Λ is small, the distances $r_{\lambda i}$ between the centers of geometry of the volumes V_λ, which enclose the grid portions λ, and the PMM atoms i can become correspondingly small and the above objective is reached. On the other hand, the computational effort will eventually increase again with decreasing R_Λ, because the number of FMM expansions grows with $\Lambda \sim (1/R_\Lambda)^3$. As a result, there will be an optimal value for R_Λ.

In summary, it is advantageous to hierarchically partition the grid charge distribution into a nested set of subdistributions of decreasing size until the optimal size is reached. From a series of preliminary DFT/PMM test calculations, we found that not more than two new hierarchy levels $l = -2$ and $l = -3$ beyond the SAMM levels[17,18] $l = -1, 0, 1, \ldots$ (cf. Sec. II B), are required (data not shown). The optimal voxel size R_Λ will be determined below in Sec. IV C.

1. Definition of a new lowest DFT/PMM cluster level

Figure 1 sketches the volume V of a DFT box by a black square, an enclosed DFT grid by $\Gamma = 64$ light gray dots, and a disjoint decomposition of V into $\Lambda = 16$ voxels λ with volumes $V_\lambda = V_\Lambda \equiv V/\Lambda$ by gray-dashed squares. The DFT box is centered around a water molecule (black) representing a sample DFT fragment. Two light gray water molecules depict a part of the PMM fragment.

The radius $R_\Lambda = V_\Lambda^{1/3}/2$ of the dotted circle inscribed into the bottom right voxel λ (gray background) is the radius of gyration of this cubic voxel. λ is surrounded by a sphere of radius $d_{\Lambda,-2} \equiv 2R_\Lambda/\Theta_\chi$ (dashed black lines), which is the smallest distance still acceptable by the FMM voxel IAC (5) for reasonably accurate FMM descriptions of $\Phi(\mathbf{r}_\gamma)$ on the grid points $\gamma \in \lambda$. For PMM atoms i closer to the geometrical center $\mathbf{r}_\lambda$ (cross) of the voxel λ, the costly electrostatic pair interactions of their Gaussian charges and dipoles with the grid charges q_γ located within λ have to be evaluated (at level $l = -3$).

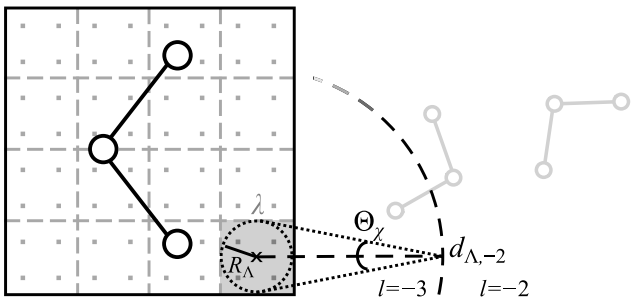


FIG. 1. Visualization of the FMM voxel IAC (5) separating the SAMM levels $l = -3$ and $l = -2$ of the DFT/PMM electrostatics computation. For explanation, see the text.

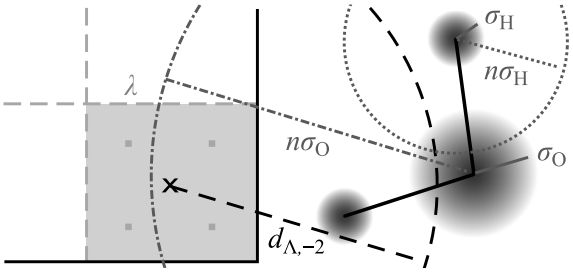


FIG. 2. Schematic illustration of the Gaussian truncation IAC (6) concentrating on the bottom right voxel λ (gray background) discussed in Fig. 1. Also here, the FMM IAC (5) is indicated by the dashed black lines. Additionally, the Gaussian IAC is depicted for an O atom by dot-dashed and for a H atom by dotted lines. The O atom apparently fails the Gaussian IAC, because it is too close to the center of geometry (cross) of λ.

If one solely considered the FMM voxel IAC (5), then the interaction of more distant PMM atoms i with the grid charges q_γ enclosed by the voxel λ would be calculated by FMM expansions, which neglect the finite Gaussian widths σ_i of the atomic charge and dipole distributions. However, at distances $r_{\gamma i}$ shorter than $n_{\text{num}}\sigma_i$, the representation of Gaussian charges and dipoles by point objects starts to entail numerical errors [$n_{\text{num}} = 8.7$ at double precision; cf. also the discussion of Eq. (2)]. If the FMM voxel IAC (5) allows such distances, these Gaussian truncation errors add up to the FMM errors.

In Sec. IV A below, we will require that the Gaussian truncation errors should be much smaller than the FMM errors. This requirement will then determine a standard value n_G for the Gaussian cutoff parameter n appearing in the additional voxel IAC,

$$r_{\lambda i} > n\,\sigma_i, \quad \text{with} \quad n < n_{\text{num}}, \tag{6}$$

according to which a Gaussian PMM source i of width σ_i can be replaced by a point-like source, if its distance $r_{\lambda i}$ from a voxel λ exceeds $n\,\sigma_i$. Thus, the FMM IAC (5) and the Gaussian truncation IAC (6) have to be simultaneously fulfilled for a sufficiently accurate FMM computation (at level $l = -2$) of the electrostatic interactions between a Gaussian PMM source i and the grid charges q_γ enclosed by the voxel λ.

Figure 2 illustrates the additional voxel IAC (6) for a PMM water molecule featuring Gaussian charge distributions of widths σ_O at the oxygen and $\sigma_H < \sigma_O$ at the hydrogen atoms, respectively. In the depicted configuration, the FMM IAC (5) allows FMM for the two rightmost atoms H and O, whereas the Gaussian truncation IAC (6) enables FMM solely for the H atom at the right. It thus enforces, in particular, the exact evaluation of the potential generated by a Gaussian charge or induced dipole of the O atom at all points γ of the voxel λ (which is the computational mode at level $l = -3$).

2. The next higher DFT/PMM cluster level

With increasing distances $r_{\lambda i}$, the $\text{SAMM}^{\chi,\text{E}}_{4,q,1}$ expansions within a voxel λ become rapidly much more accurate. At distances $r_{\lambda i} \gg 2R_\Lambda/\Theta_\chi$, a comparably accurate description can therefore be achieved by merging several portions λ of the distribution of grid charges into larger clusters.

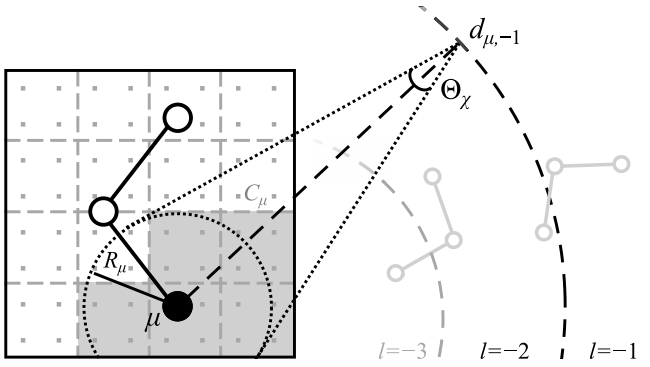


FIG. 3. Visualization of the atomic IAC (7) separating the SAMM levels $l = -2$ and $l = -1$ of the DFT/PMM electrostatics computation. For explanation, see the text.

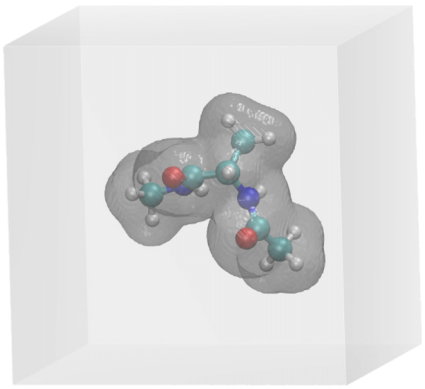


FIG. 4. An alanine dipeptide molecule in a cubic DFT box of side length 14 Å harboring the DFT grid. The gray surface encloses all those grid points, which contribute with maximal weights $w_\gamma = 1$ [cf. Eq. (8)] to the computation Eq. (10) of the radii of gyration R_μ of the DFT atoms.

Figure 3 illustrates our choice for combining voxels λ into larger atomic portions V_μ of the DFT grid. For this purpose, we employ a Voronoi tessellation of the DFT box (cf. Refs. 6 and 22), which is based on the locations $\mathbf{r}_\mu$ of the DFT atoms and assigns each voxel λ to the DFT atom μ of least distance $|\mathbf{r}_\mu - \mathbf{r}_\lambda|$ from its geometrical center $\mathbf{r}_\lambda$. Thus, each sub-volume V_μ, which belongs to a DFT atom μ, is the union $\cup V_{\lambda(\mu)}$ of all associated voxels $\lambda(\mu)$. The gray shading of five voxels in the figure exemplifies this merging of voxels for a specific DFT atom μ (black dot).

Furthermore, the figure also depicts the radius of gyration R_μ (dotted circle), which belongs to the charge distribution C_μ enclosed by V_μ, and a limiting distance $d_{\mu,-1} \equiv 2R_\mu/\Theta_\chi$ (dashed black lines). This distance follows from the properly adjusted "atomic" IAC,

$$2R_\mu/r_{\mu i} \leq \Theta_\chi, \tag{7}$$

which is analogous to the FMM voxel IAC (5). At distances $r_{\mu i} \geq d_{\mu,-1}$, FMM expansions around the position $\mathbf{r}_\mu$ of the DFT atom μ are employed to compute the electrostatic interactions of a PMM atom i with all grid charges q_γ in the cluster C_μ. At the given FMM level $l = -1$, the distances $d_{\mu,-1}$ are always large enough that the Gaussian character of the PMM charge and dipole distributions can safely be neglected.

For an atom μ at the surface of the DFT fragment, the atomic part V_μ of the DFT grid is quite large, because it extends up to the boundary of the DFT box, where the electron density vanishes by construction.[23] Then, the size of V_μ is a poor measure for the size R_μ of the atomic charge distribution C_μ. A more appropriate measure is obtained by accounting for the electron density through the positive weights

$$w_\gamma = \begin{cases} 1 & \text{for } q_\gamma > Q_e/\Gamma \\ q_\gamma/(Q_e/\Gamma) & \text{else} \end{cases}, \tag{8}$$

which enter, after atomic normalization,

$$p_\gamma^\mu \equiv \frac{w_\gamma}{\sum_{\gamma' \in V_\mu} w_{\gamma'}}, \tag{9}$$

the computation of the radii of gyration,

$$R_\mu = \left[\sum_{\gamma \in V_\mu} p_\gamma^\mu \left(\mathbf{r}_\mu - \mathbf{r}_\gamma \right)^2 \right]^{1/2}, \tag{10}$$

of the atomic charge distributions C_μ. In Eq. (8), $Q_e = \sum_\gamma q_\gamma$ is the total electron charge of the DFT fragment.

For the sample molecule alanine dipeptide, Figure 4 illustrates how the weights w_γ describe the shape and extension of an electron density. The depicted gray surface encloses all those grid points, whose grid charges q_γ exceed the threshold Q_e/Γ defined by the uniformly distributed electron density. Because of Eq. (8), these grid points carry the maximal weights $w_\gamma = 1$. According to Eqs. (9) and (10), these grid points provide the dominant contributions to the radii of gyration R_μ of the DFT atoms μ. Grid points outside the gray surface contribute only little. As a result, the R_μ actually provide estimates for the extensions of the atomic electron densities C_μ as desired for the FMM description of its electrostatic interactions with surrounding PMM atoms. In the depicted example, the R_μ are distributed around the average value of 1.19 Å with a standard deviation of only 0.15 Å. This average value is clearly in the range expected for sizes of such atoms.

The choice of the weights w_γ through Eq. (8) and the resulting radii of gyration R_μ are, of course, somewhat arbitrary. Therefore, we have studied alternative definitions entailing larger R_μ. As documented in Sec. S1 of the supplementary material,[49] we found that the above choice happens to be quite favorable.

At the highest FMM level $l = -1$ discussed so far, the atomic radii R_μ and the distances $r_{\mu i}$ enter, as we have seen above, the atomic IAC (7), which decides whether the interaction of a PMM atom i with the charge density C_μ can be treated by FMM expansions around $\mathbf{r}_\mu$. At the next higher level $l = 0$ of the SAMM cluster hierarchy,[17,18,24] atoms are combined into the predefined SUs (cf. Sec. II B). As reference points for the SAMM expansions within atomic clusters at levels $l \geq 0$ we choose the geometrical centers of the constituent atoms. If such clusters contain voluminous DFT atoms, then the computation of cluster sizes should inherit the extensions of the constituent atoms. For this purpose, Sec. S2 of the supplementary material[49] introduces for all upper cluster levels $l \geq 0$ a computationally efficient bottom-up strategy, by which one can calculate the radii of gyration of large clusters from tensorial moments of gyration characterizing the contained sub-clusters. This strategy resembles that of

computing electrostatic moments bottom-up toward the top of a FMM tree.

D. Discussion of the revised DFT/PMM cluster scheme

The above introduction of the two new cluster levels covers the central ideas underlying our adaptation of the DFT/PMM interface to the novel IAC decision strategy realized by $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$. The concepts and details of the import of the external potential $\Phi(\mathbf{r}_\gamma)$ onto points γ of the grid as well as the export of the electrostatic potential and its derivatives generated by the various portions of the DFT charge distribution at the positions $\mathbf{r}_i$ of the PMM atoms have been previously specified.[6] For completeness, these issues are presented for the revised cluster scheme in Sec. S3 of the supplementary material.[49]

At levels higher than $l = -1$, which are omitted in Fig. 3 but are indicated in the corresponding Figure 1 of the predecessor paper[6] (with the slightly different level numbering scheme $l_{\mathrm{pre}} \equiv l + 2$), the $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$ computation of the electrostatic interactions is identical to that in pure PMM systems, which has been recently described in great detail.[24]

As mentioned above in Sec. I, we have adopted with $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$ another most recent[42] algorithmic improvement. Accordingly, the FMM forces are now exact negative gradients of the FMM energy[42] and thus enable energy conserving MD simulations also for clusters moving relative to each other. Section S4 in the supplementary material[49] illustrates this fact by a sample DFT/PMM-MD simulation.

The remainder of this contribution will serve to illustrate the performance gains entailed by the consequent integration of a grid-based DFT method into the up-to-date FMM scheme $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$, which had been originally developed[24,42] for the treatment of long-range interactions in PMM-MD simulations.

III. METHODS

As our condensed-phase DFT/PMM test system we chose an alanine dipeptide molecule (Ac-Ala-NHMe) representing the DFT fragment, which is embedded in a system of PMM water models described by the recent TL6P potential.[43,50]

A. DFT/PMM simulation setup

The DFT fragment Ac-Ala-NHMe was centered in a cubic DFT box of size $(14\ \text{Å})^3$. It was described by the gradient-corrected exchange functional of Becke,[51] by the correlation functional of Lee, Yang, and Parr,[52] and by the norm-conserving pseudo-potentials of Troullier and Martins.[53] The chosen plane-wave cutoff of 100 Ry led to $\Gamma = 180^3$ grid points. The DFT convergence criterion was $\chi_{\mathrm{DFT}} = 10^{-8}$ in accuracy checks and 10^{-6} in DFT/PMM-MD simulations.

The polarizable six-point potential TL6P[43] for water features a rigid liquid phase geometry, two positive charges at the H atoms, three massless negative charges M, L_1, and L_2 in the vicinity of the O atom, and a Gaussian induced dipole distribution of width $\sigma_{\mathrm{O}} = 0.806$ Å at O. Van der Waals interactions among TL6P water molecules are represented by Buckingham potentials, whereas those with the atoms of the DFT fragment are expressed through an almost equivalent 12-6 Lennard-Jones potential.[43] Lennard-Jones parameters for alanine dipeptide were taken from the CHARMM22 force field.[39] The dielectric constant of the RF was $\epsilon = 78$ and the PMM convergence threshold was $\chi_{\mathrm{PMM}} = 10^{-4}$ D.[6]

According to Sec. II A, static partial charges of PMM atoms are represented by Gaussian distributions, if they are in the immediate neighborhood of the DFT fragment. Following Ref. 54, we chose the widths $\sigma_{\mathrm{M}} = \sigma_{\mathrm{L}} = 0.46$ Å for the in-plane (M) and out-of-plane (L_1, L_2) massless charges and $\sigma_{\mathrm{H}} = 0.24$ Å for the partial charges at the H atoms.

Simulation systems were constructed in a PMM setting. A periodic cubic box with the volume $V = (51.29\ \text{Å})^3$ was filled with $N = 4487$ TL6P water molecules and one CHARMM22 model[39] of alanine dipeptide such that the system's density corresponds to the experimental density of water at standard conditions. The system was equilibrated by $\mathrm{SAMM}^{\mathrm{f,E}}_{4,1,1}$/RF-MD for 1.5 ns with a Bussi thermostat[55] (target temperature $T_0 = 300$ K, relaxation time 1 ps) using a 1 fs time step for the dynamics integration with the velocity Verlet algorithm[56] and keeping all bond lengths involving H atoms fixed by the MSHAKE[57] and RATTLE[58] algorithms (relative tolerance 10^{-10}). From the end of this simulation, 16 statistically independent snapshots were drawn at 20 ps temporal delays.

Accuracy and efficiency of the revised DFT/PMM interface were checked for the three $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ algorithms[24] defined by $\chi \in \{\mathrm{f,m,a}\}$ and were compared with the performance of the previous[6,21] interface, which had been based on the SAMM_4 distance class scheme. The DFT/PMM accuracy parameter[6] $d(C^0_\mu)$ of SAMM_4 was chosen as 8 Å. The original suggestion[6] of 6 Å turned out to yield too poor approximations of the external potential near the edges of the DFT box, occasionally causing convergence problems of the KS orbitals. The revised choice $d(C^0_\mu) = 8$ Å removed these instabilities.

For accuracy checks, the 16 snapshots were employed to generate an ensemble $\mathcal{S}$ of static solvation structures. Taking advantage of the periodicity, the DFT fragment was shifted to the center of the PMM simulation box. The sources generating the external potential $\Phi(\mathbf{r}_\gamma)$ at the points of the DFT grid were restricted to the TL6P water molecules in the central simulation box, thus neglecting the periodic copies. This restriction served to avoid possible artifacts connected with the boundary at d_{MIC}, at which the explicit electrostatics treatment is smoothly changed into a RF description, but does not hamper the transferability of the accuracy measurements to fully periodic systems described by $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$/RF.

B. Accuracy measures

To assess the accuracies of the four SAMM algorithms, which were mentioned above and offer different FMM approximations for the computation of the DFT/PMM electrostatics, we compare, on the whole DFT grid, the deviations of the approximate external potentials Φ from associated exact references Φ^{ref}. For a most compact notation of our basic accuracy measure, we introduce the average $\langle f \rangle_\Gamma \equiv (1/\Gamma) \sum_\gamma f(\mathbf{r}_\gamma)$ of a quantity $f(\mathbf{r}_\gamma)$ defined at all Γ points γ of the DFT grid. Next, we consider for a given structural snapshot

the dimensionless root mean square deviation (RMSD)

$$\xi(\Phi,\Phi^{\mathrm{ref}}) = \sqrt{\frac{\langle(\Phi-\Phi^{\mathrm{ref}})^2\rangle_\Gamma}{\langle(\Phi^{\mathrm{ref}})^2\rangle_\Gamma - \langle\Phi^{\mathrm{ref}}\rangle_\Gamma^2}} \tag{11}$$

of an approximate potential Φ from the respective exact reference Φ^{ref}, whose variance provides an intrinsic scale. Then, our basic accuracy measure $\langle\xi(\Phi,\Phi^{\mathrm{ref}})\rangle_{\mathcal{S}}$ is the average $\langle\ldots\rangle_{\mathcal{S}} \equiv (1/16)\sum_{\mathcal{S}}\ldots$ over the snapshot ensemble $\mathcal{S}$.

The accuracy of the external $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ potential Φ, which acts on the DFT fragment, is steered by the parameters Θ and n [cf. the FMM IACs (3), (5), (7), and the Gaussian truncation IAC (6)], that is,

$$\Phi(\mathbf{r}_\gamma) \equiv \Phi(\mathbf{r}_\gamma \,|\, \Theta, n). \tag{12}$$

For the computation of the exact references Φ^{ref}, the following parameters were chosen:

$$\Phi^{\mathrm{ref}}(\mathbf{r}_\gamma) \equiv \Phi(\mathbf{r}_\gamma \,|\, 0, n_{\mathrm{num}}). \tag{13}$$

Because the threshold Θ of the FMM IACs (3), (5), and (7) is set to zero, the minimal distance between finite size clusters, which is just compatible with a FMM description, diverges. Therefore, all DFT/PMM interactions are calculated on level $l = -3$, i.e., the potential generated by the charges and dipoles of all PMM atoms is evaluated by exact pair expressions at all grid points γ. The choice $n = n_{\mathrm{num}}$ in the Gaussian truncation IAC (6) dictates that the action of the Gaussian sources is computed on the whole grid at numerical accuracy.

The FMM accuracy can be separately analyzed by considering the ensemble average deviation D_{FMM} of potentials Φ_{P} and $\Phi^{\mathrm{ref}}_{\mathrm{P}}$, which are generated by point-like sources ($\sigma_i = 0$). With the arbitrary choice $n = 0$ in the Gaussian IAC (6), this deviation,

$$D_{\mathrm{FMM}}(\Theta_\chi) \equiv \left\langle\xi[\Phi_{\mathrm{P}}(\Theta_\chi,0),\Phi^{\mathrm{ref}}_{\mathrm{P}}(0,0)]\right\rangle_{\mathcal{S}}, \tag{14}$$

solely depends on the parameter Θ_χ, which tunes the FMM IACs (3), (5), and (7) used for the computation of the approximate potential $\Phi_{\mathrm{P}}(\mathbf{r}_\gamma \mid \Theta_\chi, 0)$.

Similarly, one can single out the size of the Gaussian truncation error by considering the deviation

$$D_{\mathrm{Gauss}}(n) \equiv \left\langle\xi[\Phi(0,n),\Phi^{\mathrm{ref}}(0,n_{\mathrm{num}})]\right\rangle_{\mathcal{S}}. \tag{15}$$

Here, the approximation $\Phi(\mathbf{r}_\gamma \mid 0, n)$ differs from $\Phi^{\mathrm{ref}}(\mathbf{r}_\gamma \mid 0, n_{\mathrm{num}})$ solely by using in the IAC (6) a Gaussian truncation parameter $n < n_{\mathrm{num}}$.

Finally, the total error is measured by the ensemble average deviation

$$D_\Phi(\Theta_\chi,n) \equiv \left\langle\xi[\Phi(\Theta_\chi,n),\Phi^{\mathrm{ref}}(0,n_{\mathrm{num}})]\right\rangle_{\mathcal{S}} \tag{16}$$

of the approximate $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ potential $\Phi(\mathbf{r}_\gamma)$, which depends on the joint action of all IACs, (3), (5), (6), (7), from the exact reference $\Phi^{\mathrm{ref}}(\mathbf{r}_\gamma)$.

We furthermore measure the $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ errors of the total forces $\mathbf{f}_\mu$ acting on the DFT atoms μ by the average RMSD,

$$D_{\mathbf{f}}(\Theta_\chi,n) \equiv \frac{1}{\sigma_{\mathbf{f}}}\left\langle\sqrt{\frac{1}{3M}\sum_\mu\sum_\alpha\left[f_{\mu,\alpha}(\Theta_\chi,n) - f^{\mathrm{ref}}_{\mu,\alpha}\right]^2}\right\rangle_{\mathcal{S}} \tag{17}$$

measured in units of the average standard force deviation

$$\sigma_{\mathbf{f}} \equiv \left\langle\sqrt{\frac{1}{3M}\sum_\mu\sum_\alpha\left(f^{\mathrm{ref}}_{\mu,\alpha}\right)^2}\right\rangle_{\mathcal{S}}, \tag{18}$$

where μ counts the M DFT atoms and α the Cartesian components x, y, and z. The force components $f^{\mathrm{ref}}_{\mu,\alpha}$ are calculated by CPMD from the reference potential Φ^{ref}, whereas the forces $f_{\mu,\alpha}(\Theta_\chi,n)$ are calculated from the approximate potential $\Phi(\Theta_\chi,n)$.

Note that the force errors (17) are calculated, just as those of the potential [Eq. (16)], in a "realistic" setting, i.e., they include the FMM and the Gaussian truncation errors. Here, one cannot separately address the FMM errors, because replacing the Gaussian PMM charges and dipoles by point sources, which would be required for this error measurement [cf. Eq. (14)], almost surely prevents a convergence of the KS orbitals.

With the aim of demonstrating the progress achieved by the transition to $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ through the revised interface, we computed also with the previous SAMM_4 interface corresponding average deviations $D_x(\mathrm{SAMM}_4)$, $x \in \{\mathrm{FMM},\Phi,\mathbf{f}\}$.

For $x \in \{\Phi,\mathbf{f}\}$, these SAMM_4 data serve then as references for the accuracy gains

$$\mathcal{A}_x(\chi) \equiv \frac{1}{D_x(\Theta_\chi,n)/D_x(\mathrm{SAMM}_4)} \tag{19}$$

of the $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ algorithms relative to their predecessor.

C. Efficiency measures

For estimates of the computational efficiency, we measured the total computer time $T_{\mathbf{f}}$ consumed per integration step in 100 fs DFT/PMM-MD simulations of the dipeptide solvated in TL6P water, which were executed in the NVE ensemble with a time step of 0.5 fs. During such a DFT/PMM-MD simulation, the KS orbitals and the induced PMM dipoles have to be simultaneously brought to self-consistence. The DFT computations executed for this task usually take much longer than the time T_Φ spent for the FMM-based import and export of the electrostatics onto and from the DFT grid, which we also measured.

The sketched timings T_x, $x \in \{\Phi,\mathbf{f}\}$, were used to compute efficiency gains

$$\mathcal{E}_x(\chi) \equiv \frac{1}{T_x(\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1})/T_x(\mathrm{SAMM}_4)} \tag{20}$$

of the $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ algorithms relative to SAMM_4.

The timings for efficiency checks were obtained using $C = 16$ cores making up one node of the SuperMUC Petascale System at the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities in Munich. This system was additionally used for scalability checks of the parallel IPHIGENIE/CPMD implementation. Here, timings $T_{\mathbf{f}}(C \mid \mathrm{SAMM}^{\mathrm{a,E}}_{4,1,1})$ were measured as a function of the number C of employed cores and the parallel speedup is then given by

$$\mathcal{U}_{\mathbf{f}}(C) \equiv \frac{1}{T_{\mathbf{f}}(C \mid \mathrm{SAMM}^{\mathrm{a,E}}_{4,1,1})/T_{\mathbf{f}}(1 \mid \mathrm{SAMM}^{\mathrm{a,E}}_{4,1,1})}. \tag{21}$$

Analogously, one may define such a speedup $\mathcal{U}_{\mathrm{DFT}}(C)$ for the DFT part of the calculation.

IV. RESULTS

The new DFT/PMM interface, which was adjusted to the $\mathrm{SAMM}_{4,1,1}^{\chi,\mathrm{E}}$ family of algorithms, features two free parameters, for which suitable values remain to be chosen. One is the size R_Λ of the voxels, which disjointly partition the DFT grid into small sub-volumes as described in Sec. II C, and the other the parameter n defined by the Gaussian truncation IAC (6). Here, the Gaussian cutoff parameter n can affect the accuracy, by which the external potential $\Phi(\mathbf{r}_\gamma)$ is computed. After a choice n_G for n which guarantees that the Gaussian truncation error $D_{\mathrm{Gauss}}(n_\mathrm{G})$ [cf. Eq. (15)] is always much smaller than the FMM errors $D_{\mathrm{FMM}}(\Theta_\chi)$ [cf. Eq. (14)], this accuracy should not depend on R_Λ because, by construction of the FMM IACs (3), (5), and (7), the FMM errors $D_{\mathrm{FMM}}(\Theta_\chi)$ are generally independent of cluster sizes and, thus, particularly of the voxel size R_Λ.

Figure S20 in Sec. S5 of the supplementary material[49] demonstrates that the FMM errors $D_{\mathrm{FMM}}(\Theta_\chi)$ are actually independent of R_Λ also in our implementation. With the choice $R_\Lambda \approx 0.35$ Å, which is arbitrary at this stage of the discussion, we will now study the effects of the parameter n on the Gaussian truncation errors $D_{\mathrm{Gauss}}(n)$.

A. Acceptable Gaussian truncation errors

Using the alanine dipeptide molecule solvated in TL6P water as a typical sample system, we will now choose a default value n_G for n from the requirement that the Gaussian truncation error $D_{\mathrm{Gauss}}(n_\mathrm{G})$ should be at least by two orders of magnitude smaller than the FMM errors $D_{\mathrm{FMM}}(\Theta_\chi)$. This choice seems to be reasonable, because it most likely guarantees that the accuracy of the DFT/PMM forces, which act on the DFT atoms, almost exclusively depends on the accuracy of the FMM approximation.

Figure 5 enables such a choice, because it compares the FMM errors $D_{\mathrm{FMM}}(\Theta_\chi)$ of $\Phi(\mathbf{r}_\gamma)$ resulting for the three algorithmic variants χ with the n-dependence of the Gaussian truncation error $D_{\mathrm{Gauss}}(n)$ for $n \in [2.0, 5.7]$. The Gaussian truncation IAC (6) implies for smaller values of n that Gaussian sources (charge/dipole) are substituted by point-like objects at correspondingly smaller distances and, thus, render larger truncation errors $D_{\mathrm{Gauss}}(n)$. Taking the FMM error $D_{\mathrm{FMM}}(\Theta_\mathrm{a})$ (red dashed line) of the most accurate variant $\mathrm{SAMM}_{4,1,1}^{\mathrm{a,E}}$ as a guideline, the default $n_\mathrm{G} \equiv 4$ for n immediately follows from the accuracy requirement voiced above.

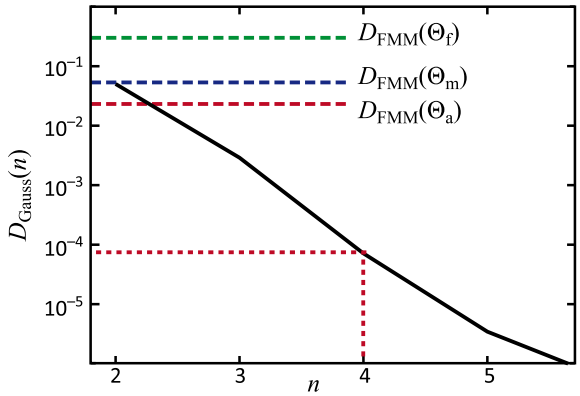


FIG. 5. Error $D_{\mathrm{Gauss}}(n)$ as a function of the Gaussian truncation parameter n defined by the IAC (6) (solid line). For comparison, the dashed horizontal lines give the FMM errors $D_{\mathrm{FMM}}(\Theta_\chi)$ for the algorithmic variants $\chi \in \{\mathrm{f}, \mathrm{m}\,\mathrm{a}\}$ (green, blue, red; cf. also Fig. S20 in the supplementary material[49]). The dotted line indicates our choice $n_\mathrm{G} \equiv 4$ for n and the corresponding $D_{\mathrm{Gauss}}(n_\mathrm{G})$.

B. Optimization of voxel sizes

The choice $n = n_\mathrm{G}$ has fixed one of the two parameters, which specify the revised DFT/PMM interface. The remaining parameter is the voxel size R_Λ, which does not affect the accuracy but should influence the computational efficiency as we have argued in the discussion of the FMM voxel IAC (5).

Equation (20) defines the efficiencies $\mathcal{E}_\Phi(\chi)$, which are gained, as compared to SAMM_4, in the DFT/PMM electrostatics computation with $\mathrm{SAMM}_{4,1,1}^{\chi,\mathrm{E}}$. Figure 6 shows how the various efficiency gains $\mathcal{E}_\Phi(\chi)$ depend on the voxel size R_Λ. According to the displayed data, the gains $\mathcal{E}_\Phi(\chi)$ become, independent of the chosen accuracy level χ, maximal for voxel sizes $R_\Lambda^{\mathrm{opt}} \approx 0.35$ Å. The existence of an optimal size is readily understood by the following consideration:

According to the FMM voxel IAC (5), an increase of R_Λ reduces the number of PMM atoms, whose interactions with the grid charges in a voxel can be efficiently treated by FMM expansions at level $l = -2$. On the other hand, a decrease of R_Λ cubically increases the number of voxels partitioning the DFT grid, and hence, a likewise increasing number of FMM expansions has to be handled. These conflicting effects are balanced at an optimal size R_Λ^{opt}, which happens to be ≈0.35 Å. Note that our program automatically partitions a given grid by voxels, whose size is close to R_Λ^{opt}. Section S6 of the supplementary material[49] generalizes this partitioning to rectangular DFT boxes.

With the choices $n_\mathrm{G} = 4$ and $R_\Lambda^{\mathrm{opt}} \approx 0.35$ Å for the Gaussian truncation parameter and for the target voxel size, respectively, the adaptation of the DFT/PMM interface to the $\mathrm{SAMM}_{4,1,1}^{\chi,\mathrm{E}}$ family of FMM algorithms is complete. The benefits of this transition can now be judged by considering the performance gains,

$$\mathcal{P}_x(\chi) \equiv \mathcal{E}_x(\chi)\mathcal{A}_x(\chi), \tag{22}$$

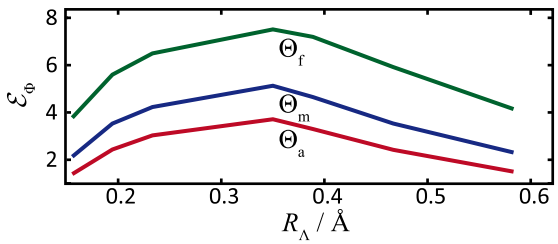


FIG. 6. Efficiency gains $\mathcal{E}_\Phi$ of the import and export of the electrostatic interactions onto and from the DFT grid [cf. Eq. (20)] measured for $\mathrm{SAMM}_{4,1,1}^{\chi,\mathrm{E}}$ ($\chi \in \{\mathrm{f}, \mathrm{m}, \mathrm{a}\}$, green, blue, red) relative to SAMM_4 as a function of the voxel size R_Λ.

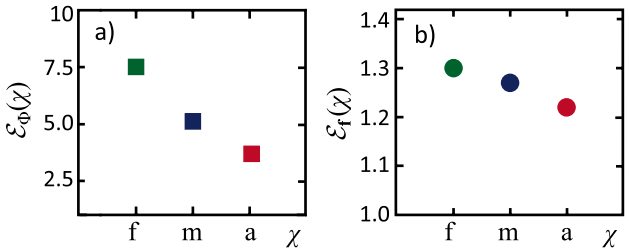


FIG. 7. $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ efficiency gains $\mathcal{E}_x(\chi)$ relative to SAMM_4 measured for the computation of (a) the electrostatic interactions between the DFT and PMM fragments per integration step ($x = \Phi$) and (b) an entire DFT/PMM-MD integration step ($x = \mathbf{f}$).

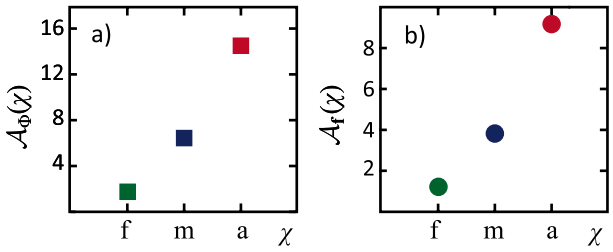


FIG. 8. $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ accuracy gains $\mathcal{A}_x(\chi)$ relative to SAMM_4 resulting for (a) the potential $\Phi(\mathbf{r}_\gamma)$ on the DFT grid ($x = \Phi$) and (b) the forces $\mathbf{f}_\mu$ acting on the DFT atoms ($x = \mathbf{f}$).

relative to the previous SAMM_4-based interface with respect to two tasks x, i.e., the computation of the electrostatic interactions between the two fragments ($x = \Phi$) and the computation of the forces acting on the DFT atoms ($x = \mathbf{f}$). Equation (22) defines the performance gains for these tasks simply as the product of the efficiency and accuracy gains $\mathcal{E}_x(\chi)$ [Eq. (20)] and $\mathcal{A}_x(\chi)$ [Eq. (19)], respectively. Note that, in general, accuracy and efficiency are mutually contradicting aspects of algorithmic performance. We start the discussion of the achieved progress by considering the efficiency gains $\mathcal{E}_x(\chi)$.

C. Efficiency gains

The fact that the revised DFT/PMM interface handles the subtask $x = \Phi$ more efficiently than its predecessor becomes immediately apparent, if one looks once again at Figure 6 discussed above. The figure shows for the optimal voxel size R^{opt}_Λ the maximal efficiency gains $\mathcal{E}_\Phi(\chi) = 7.5$, 5.1, and 3.7 for $\chi = \mathrm{f}$, m, and a, respectively.

These maximal gains $\mathcal{E}_\Phi(\chi)$ are once more displayed by Figure 7(a) to enable a visual comparison with the efficiency gains $\mathcal{E}_\mathbf{f}(\chi)$ of the complete force calculation per DFT/PMM-MD integration step, which are shown in Figure 7(b). The total efficiency gains $\mathcal{E}_\mathbf{f}(\chi)$ additionally cover the times consumed by the self-consistent computations of the KS orbitals and of the induced PMM dipoles. They are in the range between 30% (f), 27% (m), and 22% (a) and, thus, are much smaller than the gains $\mathcal{E}_\Phi(\chi)$ of the DFT/PMM electrostatics subtask. The latter gains do not transfer into correspondingly large total gains $\mathcal{E}_\mathbf{f}(\chi)$, because the electrostatics subtask consumes only 3% (f) up to 5% (a) of the total time. Nevertheless, all $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ algorithms are noticeably more efficient than SAMM_4.

D. Accuracy gains

Now, the general contradiction between efficiency and accuracy raises the question, whether the noted efficiency gains are accompanied by accuracy losses. In contrast to this simple conjecture, Figure 8 demonstrates that the revision of the DFT/PMM interface leads, for all three $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ variants, to sizable accuracy gains $\mathcal{A}_x(\chi)$ as defined by Eq. (19) for the computations of (a) the external potential $\Phi(\mathbf{r}_\gamma)$ and of (b) the forces acting on the DFT atoms.

The comparison of Figures 8(a) and 8(b) demonstrates that the accuracy gains $\mathcal{A}_\Phi(\chi)$ achieved for the computation of the external potential $\Phi(\mathbf{r}_\gamma)$ are for all variants χ by about a factor of 1.5 larger than the gains $\mathcal{A}_\mathbf{f}(\chi)$ for the force computation. Thus, the enhanced accuracy of the external potential is not completely transferred through the self-consistent computation of the KS orbitals and of the induced PMM dipoles to the forces on the DFT atoms. Nevertheless, all $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ forces $\mathbf{f}_\mu$ are more accurate than those calculated with SAMM_4 and this advantage amounts to about one order of magnitude for the most accurate variant $\chi = \mathrm{a}$ (Fig. 8(b)).

Within the new class of DFT/PMM algorithms, however, accuracy gains of a variant χ should be measured with respect to the most inaccurate variant $\chi = \mathrm{f}$. For $\chi \in \{\mathrm{m}, \mathrm{a}\}$, the data in Figure 8 then yield almost identical relative accuracy gains concerning the potential $\Phi(\mathbf{r}_\gamma)$ and the atomic forces $\mathbf{f}_\mu$, i.e., one finds

$$\frac{\mathcal{A}_\Phi(\chi)}{\mathcal{A}_\Phi(\mathrm{f})} \approx \frac{\mathcal{A}_\mathbf{f}(\chi)}{\mathcal{A}_\mathbf{f}(\mathrm{f})} \quad \text{for} \quad \chi \in \{\mathrm{m}, \mathrm{a}\}. \tag{23}$$

Thus, more accurate potentials, which are obtained within the new class of $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ interface algorithms by using tighter FMM thresholds $\Theta_\chi < \Theta_\mathrm{f}$ in the IACs (3), (5), and (7), directly translate into more accurate forces. The relative gains $\mathcal{A}_\mathbf{f}(\chi)/\mathcal{A}_\mathbf{f}(\mathrm{f})$ amount to a factor of about 3 for $\chi = \mathrm{m}$ and about 7.5 for $\chi = \mathrm{a}$.

The direct translation of the more accurate potentials $\Phi(\mathbf{r}_\gamma)$ into more accurate forces $\mathbf{f}_\mu$, which is expressed by Eq. (23), additionally demonstrates that the accuracy gains of the forces shown in Figure 8(b) are almost exclusively due to the revised DFT/PMM electrostatics scheme presented in Sec. II. The added FMM expansions for the Lennard-Jones interactions and the newly introduced energy corrections for the forces,[42] in contrast, are of negligible importance for the shown DFT/PMM accuracy gains $\mathcal{A}_\mathbf{f}(\chi)$ (data of corresponding test calculations not shown).

E. Total performance gains

Because the adaptation of the DFT/PMM interface to the new class of $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ algorithms[24,42] has rendered sizable gains of accuracy and efficiency, it leads, by Eq. (22), to a considerably improved performance.

A comparison of Figures 9(a) and 9(b) shows that the performance gains $\mathcal{P}_\Phi(\chi)$ reached for the subtask of the DFT/PMM electrostatics computation are by factors 8 ($\chi = \mathrm{f}$)

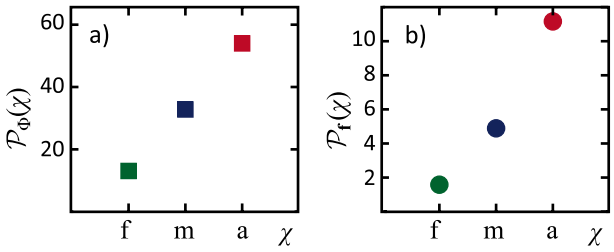


FIG. 9. Total $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ performance gains $\mathcal{P}_x(\chi) \equiv \mathcal{E}_x(\chi)\mathcal{A}_x(\chi)$ compared to SAMM_4 calculated by Eq. (22) from the efficiency and accuracy gains depicted in Figures 7 and 8.

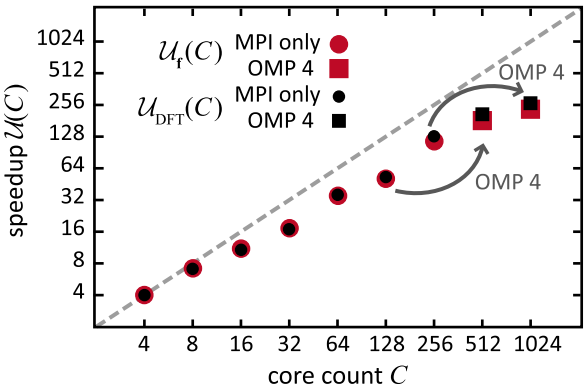


FIG. 10. Speedups $\mathcal{U}_{\mathbf{f}}(C)$ of a DFT/PMM-MD integration step (red) using C cores are compared with corresponding speedups $\mathcal{U}_{\mathrm{DFT}}(C)$ of the DFT subtask (black). Circles refer to a pure MPI setup, whereas squares refer to a hybrid parallelization employing four OpenMP threads (OMP 4) per MPI process, whose number then is $C/4$. The gray dashed line marks the ideal linear scaling.

to 5 (χ = a) larger than the total gains $\mathcal{P}_{\mathbf{f}}(\chi)$. All $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ algorithms outperform their predecessor SAMM_4 in both respects. The performance figures resemble the corresponding accuracy graphs, 8(a) and 8(b), indicating that our algorithmic revision of the DFT/PMM interface was focused on enhanced accuracy rather than efficiency.

Figure 9(b), in particular, demonstrates that the total DFT/PMM performance of $\mathrm{SAMM}^{\mathrm{a,E}}_{4,1,1}$ beats that of SAMM_4 by about one order of magnitude, whereas the gains of total performance are smaller for the other variants $\chi \in \{\mathrm{f}, \mathrm{m}\}$. Hence, $\mathrm{SAMM}^{\mathrm{a,E}}_{4,1,1}$ is apparently the best choice for DFT/PMM-MD simulations with the IPHIGENIE/CPMD program package, because its accuracy is better by factors 7.5 and 2.4, respectively, than that of the other variants $\chi \in \{\mathrm{f}, \mathrm{m}\}$, whereas its efficiency is only by 7% and 3% smaller (cf. Figs. 8(b) and 7(b)). Note that these conclusions equally well apply to the $\mathrm{SAMM}^{\mathrm{a,E}}_{4,3,1}$ algorithm, which applies a more accurate FMM expansion to the dispersion attraction, because the associated increase of computational effort is negligible in a DFT/PMM setting.

F. Scalability of the MPI/OpenMP-parallel implementation

Despite the considerable progress achieved by the integration of the electrostatic interactions between the DFT and PMM fragments of a simulation system into a multiple-scale FMM setting, DFT/PMM-MD simulations of large DFT fragments embedded in extended PMM condensed phase environments still pose huge computational tasks. Taking advantage of parallel processing on many-core computers is mandatory, if one wants to make such tasks manageable.

Therefore, we have spent some effort to jointly parallelize the combination of the program packages IPHIGENIE (PMM-MD) and CPMD (DFT) in such a way that it becomes suitable for HPC systems. For this purpose, we have thoroughly revised the technical aspects of the implementation.

Here, the FORTRAN code of CPMD is compiled as a library and linked to the C program IPHIGENIE into one executable, which allows a smooth and robust startup. During the startup phase, the MD program IPHIGENIE calls the FORTRAN routines of CPMD to set up the DFT calculation. In the course of a simulation, CPMD is invoked at each integration step through subroutine calls and manages the exchange of the necessary data with IPHIGENIE by calling C routines. Different numbers of MPI processes and/or OpenMP threads can be assigned to the PMM and DFT calculations, respectively, allowing one to independently tune optimal parallel setups for each of the two cooperating programs.

Although our sample DFT/PMM system is relatively small for HPC applications, which are feasible with a machine like the SuperMUC addressed in Sec. III C, we have nevertheless employed it to measure the total speedup $\mathcal{U}_{\mathbf{f}}(C)$ of parallel processing gained with $\mathrm{SAMM}^{\mathrm{a,E}}_{4,1,1}$ by using C CPU cores. This speedup is defined by Eq. (21) and is compared in Figure 10 with the corresponding speedup $\mathcal{U}_{\mathrm{DFT}}(C)$ achieved by the well-scaling program CPMD for the DFT subtask.

It is the most important message of Figure 10 that the total speedups $\mathcal{U}_{\mathbf{f}}(C)$ of IPHIGENIE/CPMD are almost identical to the DFT speedups $\mathcal{U}_{\mathrm{DFT}}(C)$ of CPMD for all core numbers C in the considered range. This is a consequence of the fact that IPHIGENIE and the DFT/PMM interface parallelize sufficiently well such that the computational effort required for the DFT subtask still dominates the total effort at large C. As a result, one can execute, e.g., a 30 ps DFT/PMM trajectory of our sample simulation system within one day, such that three independent trajectories suffice[59] to obtain a well-resolved infrared spectrum of alanine dipeptide solvated in water by Fourier transforming the trajectory $\boldsymbol{\mu}(t)$ of its DFT dipole moment.

V. SUMMARY

We have adjusted the DFT/PMM interface,[6] which integrates the grid-based DFT program CPMD into the PMM-MD simulation program IPHIGENIE, to the recently suggested[24,42] class of hierarchical FMM algorithms called $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$, which represents the FMM forces as negative gradients of the FMM energy and replaces the previous cutoff of the 12-6 Lennard-Jones interactions by FMM expansions of q'th order for the dispersive and of first order for the repulsive contributions.

For the purpose of this adjustment, we have partitioned the electron density of the DFT fragment, which is represented within the DFT box of volume V by the grid charges q_γ, into hierarchically nested charge clusters. The clusters at the lowest level of this hierarchy ($l = -3$) are simply the grid charges q_γ.

One level above ($l = -2$) these charges are the contents of equally sized voxels $V_\lambda \subset V$, whose radius of gyration is about 0.35 Å. At the next higher level ($l = -1$) of the DFT charge cluster hierarchy, the voxels V_λ are associated to the DFT atoms μ by the minimum distance criterion and are merged into atomic portions V_μ of the box volume V. These partial volumes V_μ harbor the atomic charge distributions C_μ with radii of gyration [Eq. (10)] of about 1.2 Å. The characterized DFT atoms are combined into predefined SUs, adaptively formed clusters of SUs etc. making up the SAMM object hierarchy[24] on levels $l \geq 0$. The DFT atoms hand their sizes up to these higher levels by the procedures described in Sec. S2 of the supplementary material.[49]

Given the hierarchically nested levels of (DFT) charge distributions sketched above, the revised interface simply applies the top-down FMM approximation strategy characteristic for $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$. Starting at the highest level of the object hierarchy, which harbors the largest clusters, the (average) apparent cluster sizes ϑ are compared with the threshold size Θ_χ to decide whether the non-bonded interactions can be approximated at the given hierarchy level l by FMM expansions. The corresponding FMM IACs employed for these decisions are given by Eqs. (3), (5), and (7).

At the lowest level ($l = -3$), the DFT grid charges q_γ interact with charges and induced dipoles of PMM atoms i, which are represented as Gaussian distributions of widths σ_i to avoid artificial perturbations of the electron density. Already at the next higher level ($l = -2$), these distributions are replaced by point objects, and the external potential $\Phi(\mathbf{r}_\gamma)$ generated by the point charges and induced point dipoles within the voxel V_λ is computed by FMM. To keep the associated Gaussian truncation error by two orders of magnitude smaller than the FMM errors, the additional IAC $r_{\lambda i} > 4\,\sigma_i$ has to hold at level $l = -2$ [cf. Eq. (6)].

For the most accurate $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ variant defined by $\chi = \mathrm{a}$ (i.e., by the FMM threshold $\Theta_\mathrm{a} = 0.17$), we obtained for the revised DFT/PMM interface a performance which beats that of its predecessor by one order of magnitude as measured for a sample DFT/PMM simulation system comprising 22 435 charges and 4487 induced dipoles in the aqueous PMM environment and 22 atoms in the DFT model of the solute alanine dipeptide. Here, the forces on the DFT atoms are by about the same factor more accurate than previously, although also the overall computational speed has been enhanced by 22%. This speedup is close to the limit, because the revised DFT/PMM interface and the PMM part of the computation meanwhile consume less than 8% of the total computation time even for the given quite large PMM fragment, whereas the much smaller DFT part takes all the rest.

As long as the Born-Oppenheimer MD simulation of the DFT fragment is manageable with CPMD on a HPC machine, also a DFT/PMM-MD simulation of a hybrid system remains manageable with IPHIGENIE/CPMD, because both programs are now integrated into a single parallelized program, in which CPMD is the bottleneck. The previous DFT/PMM interface, in contrast, had none of the technical features required for the parallel use of many cores on a HPC system. Thus, IPHIGENIE/CPMD is now ready for large-scale and stable MD applications.

IPHIGENIE will be published in the near future (Mathias *et al.*, in preparation). But already, now the program is available for download[26] together with a preliminary documentation and several sample simulation systems. Those parts of the DFT/PMM interface, which are integrated in CPMD, are available online as a patch to the current[23] CPMD version 3.17.1 and will be contained in the next forthcoming distribution of CPMD.

We would like to note that the presented interface between the PMM-MD driver IPHIGENIE[26] and the grid-based DFT program CPMD[23] should be transferable to other grid-based DFT programs at a manageable programming effort and without posing conceptual difficulties. Moreover, such a transfer, specifically to the DFT program QUICKSTEP,[60] offered by the program suite CP2K[61] would be highly desirable, because it treats hybrid functionals like B3LYP[62,63] much more efficiently[64,65] than CPMD.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (No. SFB749/C4) and by the Bavarian Competence Network for Technical and Scientific High Performance Computing (KONWIHR-III).

[1]A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[2]H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[3]B. Mennucci, Phys. Chem. Chem. Phys. **15**, 6583 (2013).
[4]P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[5]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[6]M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan, and G. Mathias, J. Chem. Phys. **138**, 244103 (2013).
[7]R. A. Bryce, M. A. Vincent, N. O. J. Malcolm, I. H. Hillier, and N. A. Burton, J. Chem. Phys. **109**, 3077 (1998).
[8]M. Dupuis, M. Aida, Y. Kawashima, and K. Hirao, J. Chem. Phys. **117**, 1242 (2002).
[9]Z. Lu and Y. Zhang, J. Chem. Theory Comput. **4**, 1237 (2008).
[10]P. K. Biswas and V. Gogonea, J. Chem. Phys. **129**, 154108 (2005).
[11]F. Lipparini, C. Cappelli, G. Scalmani, N. De Mitri, and V. Barone, J. Chem. Theory Comput. **8**, 4270 (2012).
[12]E. Boulanger and W. Thiel, J. Chem. Theory Comput. **8**, 4527 (2012).
[13]S. Caprasecca, S. Jurinovich, L. Viani, C. Curutchet, and B. Mennucci, J. Chem. Theory Comput. **10**, 1588 (2014).
[14]A. A. Appel, SIAM J. Sci. Stat. Comput. **6**, 85 (1985).
[15]J. Barnes and P. Hut, Nature **324**, 446 (1986).
[16]L. Greengard and V. Rokhlin, J. Comput. Phys. **73**, 325 (1987).
[17]C. Niedermeier and P. Tavan, J. Chem. Phys. **101**, 734 (1994).
[18]C. Niedermeier and P. Tavan, Mol. Simul. **17**, 57 (1996).
[19]M. Eichinger, H. Grubmüller, H. Heller, and P. Tavan, J. Comput. Chem. **18**, 1729 (1997).
[20]G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[21]K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, and P. Tavan, J. Chem. Theory Comput. **8**, 3628 (2012).
[22]M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello, J. Chem. Phys. **110**, 10452 (1999).
[23]J. Hutter, A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman, and M. Parrinello, *CPMD: Car–Parinello Molecular Dynamics*, version 3.17.1, IBM Corp 1990–2008 and MPI für Festkörperforschung Stuttgart 1997–2001, www.cpmd.org.
[24]K. Lorenzen, C. Wichmann, and P. Tavan, J. Chem. Theory Comput. **10**, 3244 (2014).
[25]S. Bauer, P. Tavan, and G. Mathias, J. Chem. Phys. **140**, 104103 (2014).
[26]IPHIGENIE is available for download free of charge under the GPL licence at http://sourceforge.net/projects/iphigenie.
[27]M. P. Allen and D. Tildesley, *Computer Simulations of Liquids* (Clarendon, Oxford, 1987), Chap. 1.5.2-3, pp. 24–29.

[28]T. A. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[29]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[30]B. A. Luty, I. G. Tironi, and W. F. van Gunsteren, J. Chem. Phys. **103**, 3014 (1995).
[31]G. A. Cisneros, M. Karttunen, P. Ren, and C. Sagui, Chem. Rev. **114**, 779 (2013).
[32]J. Gao and C. Alhambra, J. Chem. Phys. **107**, 1212 (1997).
[33]K. Nam, J. Gao, and D. M. York, J. Chem. Theory Comput. **1**, 2 (2005).
[34]D. Riccardi, P. Schaefer, and Q. Cui, J. Phys. Chem. B **109**, 17715 (2005).
[35]Z. C. Holden, R. M. Richard, and J. M. Herbert, J. Chem. Phys. **139**, 244108 (2013).
[36]C. Sagui, L. G. Pedersen, and T. A. Darden, J. Chem. Phys. **120**, 73 (2004).
[37]W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[38]W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[39]A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[40]C. Oostenbrink, A. Villa, A. Mark, and W. Van Gunsteren, J. Comput. Chem. **25**, 1656 (2004).
[41]Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, J. Chem. Theory Comput. **9**, 4046 (2013).
[42]K. Lorenzen, G. Mathias, and P. Tavan "Linearly scaling and almost Hamiltonian dielectric continuum molecular dynamics simulations through fast multipole expansions" (unpublished).
[43]P. Tröster, K. Lorenzen, and P. Tavan, J. Phys. Chem. B **118**, 1589 (2014).
[44]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[45]A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
[46]T. Laino, F. Mohamed, A. Laio, and M. Parrinello, J. Chem. Theory Comput. **1**, 1176 (2005).
[47]P. K. Biswas and V. Gogonea, J. Chem. Phys. **123**, 164114 (2005).
[48]Note that the Gaussian widths of the PMM charges and induced dipoles are critical parameters of the DFT/PMM interface. They steer the strengths of the electrostatic interactions between the DFT fragment and its immediate PMM environment. Their choice should be guided by sample DFT/PMM simulations and comparisons with experimental data or reference simulations.
[49]See supplementary material at http://dx.doi.org/10.1063/1.4914329 which provides on 15 pages in six sections a total of ten figures (S11 - S20) and 13 equations (S24 - S36) as additional material to the main text. Alternative choices for the weights w_γ [Eq. (8)] are investigated (S1), and the concepts for the hierarchical computation of gyration moments (S2) are outlined. The computational scheme of electrostatic DFT/PMM interactions (S3) is discussed, energy conserving properties of the DFT/PMM interface are checked (S4) and the claim that FMM errors D_{FMM} are independent of the voxel size R_Λ (S5) is proven. Eventually the general case of rectangular voxels is discussed (S6).
[50]P. Tröster and P. Tavan, J. Phys. Chem. Lett. **5**, 138 (2014).
[51]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[52]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[53]N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[54]P. Tröster, K. Lorenzen, M. Schwörer, and P. Tavan, J. Phys. Chem. B **117**, 9486 (2013).
[55]G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[56]W. C. Swope and H. C. Anderson, J. Chem. Phys. **76**, 637 (1982).
[57]V. Kräutler, W. F. van Gunsteren, and P. Hünenberger, J. Comput. Chem. **22**, 501 (2001).
[58]H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).
[59]V. Schultheis, R. Reichold, B. Schropp, and P. Tavan, J. Phys. Chem. B **112**, 12217 (2008).
[60]J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).
[61]CP2K is available for download free of charge under the GPL licence at http://www.cp2k.org
[62]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[63]P. Stephens, F. Devlin, C. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[64]M. Guidon, F. Schiffmann, J. Hutter, and J. VandeVondele, J. Chem. Phys. **128**, 214104 (2008).
[65]M. Guidon, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **5**, 3010 (2009).

Der folgende Abdruck

Supplemental Material to:
Utilizing Fast Multipole Expansions for Efficient and Accurate
Quantum-Classical Molecular Dynamics Simulations

Magnus Schwörer, Konstantin Lorenzen, Gerald Mathias und Paul Tavan
J. Chem. Phys. **142**, 104108 (2015)

enthält zusätzliche Informationen zum vorangestellen Haupttext. In Abschnitt S1 wird eine konservativere Wahl der bei der Berechnung der Gyrationsradien eingesetzten Gewichtsparameter untersucht, Abschnitt S2 beschreibt detailliert die effiziente hierarchische Berechnung von Gyrationsradien. In Abschnitt S3 wird die Behandlung der elektrostatischen DFT/PMM-Wechselwirkungen für das neue Voxel-basierte Kopplungsschema dokumentiert, und in Abschnitt S4 die durch Anwendung des nochmals überarbeiteten SAMM$^{\chi}$-Algorithmus [127] verbesserte Energieerhaltung bewiesen. Abschnitt S5 zeigt, dass der FMM-Fehler durch die Konstruktion des Verfahrens unabhängig von der Größe der Voxel ist. In Abschnitt S6 wird schließlich der allgemeine Fall nicht-kubischer quaderförmiger Voxel diskutiert.

Supplemental Material to:

Utilizing Fast Multipole Expansions for Efficient and Accurate Quantum-Classical Molecular Dynamics Simulations

Magnus Schwörer, Konstantin Lorenzen, Gerald Mathias, and Paul Tavan[a)]

Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

[a)]Electronic mail: paul.tavan@physik.uni-muenchen.de

S1. ALTERNATIVE WEIGHTS

In Section II C 2 of the main text we have introduced weights w_γ for the grid points γ by comparing the grid charges q_γ, which represent the electron density of the given molecule, with the constant reference grid charge Q_e/Γ representing the total electronic charge Q_e uniformly distributed on the Γ grid points. According to Eq. (8) all grid points with $q_\gamma > Q_\mathrm{e}/\Gamma$ get the weight $w_\gamma = 1$ defining the interior region of the electron charge distribution $\rho_\mathrm{e}(\mathbf{r})$. The remaining grid points get the weights $w_\gamma = q_\gamma/(Q_\mathrm{e}/\Gamma)$, which rapidly decay with increasing distance from the surface surrounding the interior region of $\rho_\mathrm{e}(\mathbf{r})$.

The specific choice of the surface separating the interior and exterior regions of $\rho_\mathrm{e}(\mathbf{r})$ accomplished by Eq. (8) is somewhat arbitrary, of course. Hence, the radii of gyration (10) of the DFT atoms, which depend through the normalized atomic weights p_γ^μ defined by Eq. (9) on the weights w_γ, inherit this arbitrariness. Consequently also the algorithmic properties mediated by the atomic IAC (7) are affected.

With the aim of checking the algorithmic consequences of our choice (8) we additionally considered other definitions of the separating surface, which are defined as an one-parameter manifold by the modified weights

$$\tilde{w}_\gamma(c) = \begin{cases} 1 & \text{for } q_\gamma > cQ_\mathrm{e}/\Gamma \\ q_\gamma/(cQ_\mathrm{e}/\Gamma) & \text{else} \end{cases} . \tag{S24}$$

Obviously, one has $\tilde{w}_\gamma(1) = w_\gamma$. Choices $c < 1$ for the scaling parameter c lead to a larger volume of the interior space until it eventually, in the limit $c \to 0$, fills the whole DFT box. As a result also the atomic radii of gyration R_μ should increase with decreasing c.

To provide an example, a comparison of Figure S11, which employs the value $c = 1/5000$ to define the interior space of alanine dipeptide by $\tilde{w}_\gamma = 1$, with Fig. 4 in the main text, which is based on $c = 1$, demonstrates that the volume of the interior space actually increases with decreasing c. Next, Figure S12 illustrates the effects of this increase for choices c from the range $[10^{-8}, 1]$ on the radii of gyration R_μ of the three DFT atoms C, H_1, and H_2, which are labeled in Fig. S11 .

Atom C is surrounded by other covalently linked atoms. Therefore its tessellation volume V_C is completely inside the gray surface shown in Fig. S11. Because the electron density in V_C is large, this volume will carry relatively large weights $w_\gamma(c)$ for all choices $c < 1$ and the associated radius of gyration R_C is independent of c. In contrast, the atoms H_1 and H_2

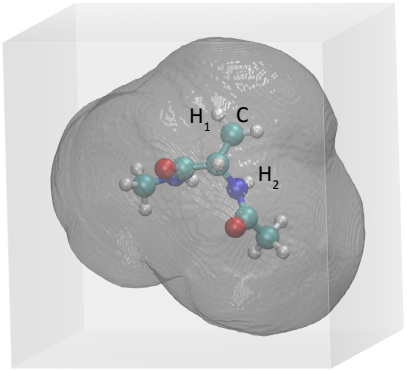


FIG. S11. An alanine dipeptide molecule in a cubic DFT box of side length 14 Å harboring the DFT grid. The gray surface encloses all those grid points, which contribute with maximal weights $\tilde{w}_\gamma(1/5000) = 1$ to the computation [Eqs. (9) and (10)] of the radii of gyration R_μ of the DFT atoms. As compared to Fig. 4 in the main text, which shows the corresponding surface for $\tilde{w}_\gamma(1)$, the interior region of $\rho_\mathrm{e}(\mathbf{r})$ has become larger.

are localized near the molecular surface such that their tessellation volumes contain many grid points outside the gray surface. Thus, changing this surface by variation of c will alter many of the weights associated to H_1 and H_2 and, therefore, the respective radii of gyration. According to Figure S12 these radii monotonously decrease with increasing $c < 1$ starting from maximal values at $c = c_0 \equiv 10^{-8}$, which lead to $w_\gamma(c_0) = 1$ for all γ.

For increasing c and correspondingly decreasing radii of gyration R_μ the limiting distance $d_{\mu,-1}$ dictated by the atomic IAC (7) becomes smaller and the potential generated by more and more nearby PMM atoms i can be efficiently covered by the FMM expansions at the atomic level $l = -1$. Thus, the accuracy of $\Phi(\mathbf{r}_\gamma)$ should conversely decrease with growing c. Using the error $D_\Phi(\Theta_\chi)$ defined by Eq. (16) as a measure, Figure S13 demonstrates that this is actually the case for all three $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ variants ($\chi \in \{\mathrm{f}, \mathrm{m}, \mathrm{a}\}$). The figure additionally shows that the errors $D_\Phi(\Theta_\chi)$ of the fastest variant $\chi = \mathrm{f}$ are almost everywhere by a factor

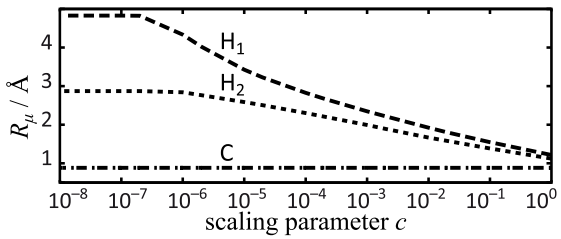


FIG. S12. Radii of gyration R_μ for three DFT atoms C, H_1, and H_2 as a function of the scaling parameter c.

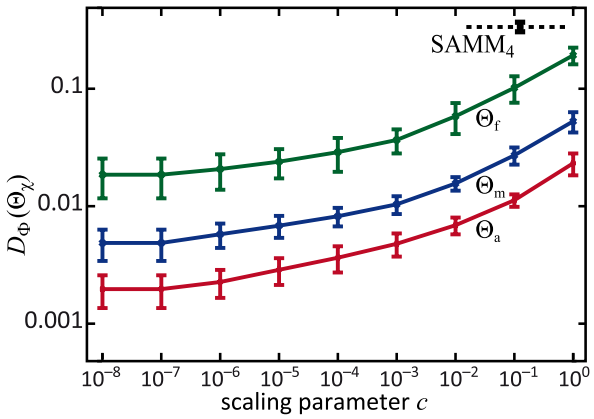


FIG. S13. The potential errors $D_\Phi(\Theta_\chi)$ as functions of c for the three $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ variants ($\chi \in \{\mathrm{f},\mathrm{m},\mathrm{a}\}$, green, blue, red). The dashed black line indicates $D_\Phi(\mathrm{SAMM}_4)$ of the predecessor algorithm. Error bars represent standard deviations.

of about 3.5 larger than those for $\chi = \mathrm{m}$ and about 10 times larger than those for $\chi = \mathrm{a}$.

The potential error $D_\Phi(\mathrm{SAMM}_4)$ of the previous SAMM_4 interface (cf. Sec. III A) is represented in Figure S13 by a black dashed line. For all χ and almost all choices of the scaling parameter c this error is much larger than the errors $D_\Phi(\Theta_\chi)$. Even at $c = 1$ it is still slightly larger than the error $D_\Phi(\Theta_\mathrm{f})$ of the most inaccurate and supposedly most efficient variant $\mathrm{SAMM}^{\mathrm{f,E}}_{4,1,1}$. Because the previous SAMM_4 based DFT/PMM interface had been accurate enough for stable DFT/PMM-MD simulations, the choice $c = 1$ seems to be reasonable. It implies quite small radii of gyration R_μ for the DFT atoms, which are favorable for an enhanced efficiency.

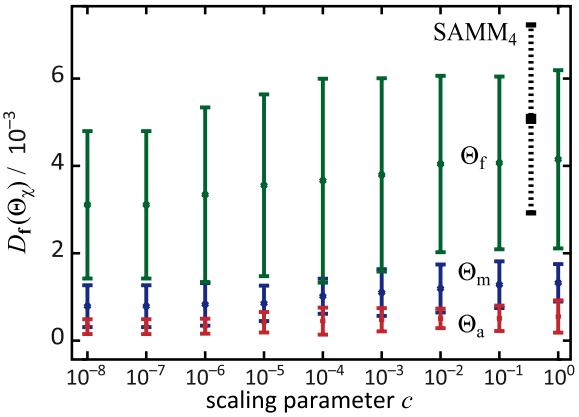


FIG. S14. The errors $D_\mathrm{f}(\Theta_\chi)$ of the forces on the DFT atoms, as functions of c for $\chi \in \{\mathrm{f},\mathrm{m},\mathrm{a}\}$ (green, blue, red) and the corresponding error for the predecessor algorithm (black).

According to Figure S14 also the errors $D_{\mathbf{f}}(\Theta_\chi)$ of the forces on the DFT atoms, which are defined by Eq. (17), increase with the scaling parameter c. Varying c from c_0 to one, the increases of the force errors amount at most to a factor of about 1.7. Thus, these increases are by a factor of about six smaller than those of the errors $D_\Phi(\Theta_\chi)$ in Fig. S13.

Summarizing we may state that the choice (8) for the sizes of the DFT atoms, which corresponds to $c = 1$ in Eq. (S24), guarantees that forces on the DFT atoms are for $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ at least as accurate as those obtained with SAMM_4. The correspondingly small R_μ render, without significant loss of accuracy (cf. Fig. S14), more efficient algorithms than the larger R_μ following from $c \ll 1$ (data on the efficiency $\mathcal{E}_{\mathbf{f}}$ as a function of c not shown).

S2. HIERARCHICAL CALCULATION OF RADII OF GYRATION

As explained in Sections I and II of the main text, the $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,1,1}$ algorithms hierarchically check in a top-down fashion by the IAC (3), whether the average aspect ratio $(\vartheta_{A,l} + \vartheta_{B,l})/2$ of a pair of clusters A and B is small enough to allow a computation of interactions on a given cluster level l. According to Eq. (4) the aspect ratio $\vartheta_{A,l}(r)$ of a cluster A linearly depends on its radius of gyration $R_{A,l}$.

Here we now will show that $R_{A,l}$ is determined by gyration tensors $\mathbf{G}^{(n)}_{A,\mathbf{0}}$ of order $n = 2$, which can be efficiently calculated in a bottom-up fashion, just like electrostatic or dispersion multipole moments,[1,2] from the gyration moments $\mathbf{G}^{(n)}_{a,\mathbf{0}}$ of its children $a \in A$.

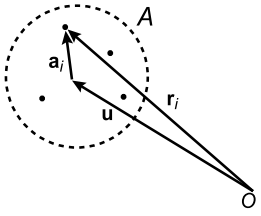


FIG. S15. Cluster geometry and vector notation.

Consider, as shown in Figure S15, a cluster A consisting of N_A objects i at the positions $\mathbf{r}_i$, an arbitrary reference point $\mathbf{u}$, and the local position vectors $\mathbf{a}_i = \mathbf{r}_i - \mathbf{u}$. Furthermore assume that the points $\mathbf{r}_i$ carry statistical weights $1 \geq p_i \geq 0$ with $\sum_i p_i = 1$, which derive from certain unnormalized weights $w_i \geq 0$ by $p_i = w_i/W_A$ with the norm

$$W_A = \sum_i w_i. \tag{S25}$$

In Section II C 2 of the main text we have employed analogous weights w_γ and their normalized counterparts p_γ^μ for the points γ of the DFT grid within the atomic volume V_μ to define by Eqs. (8)-(10) the radius of gyration for a DFT atom μ. The following general discussion contains this special case for $i \equiv \gamma$ and $p_i \equiv p_\gamma^\mu$.

For the hierarchical bottom-up computation of the radii of gyration we introduce the local gyration moments

$$\mathbf{G}_{A,\mathbf{u}}^{(n)} \equiv \sum_{i=1}^{N_A} p_i \, \mathbf{a}_i^{(n)}, \tag{S26}$$

where the symbol $\mathbf{a}_i^{(n)}$ is the n-fold outer tensor product of $\mathbf{a}_i$ with itself (Ref. 1 thoroughly explains the employed tensorial notation). Then the radius of gyration

$$R_\mathbf{u} = \sqrt{\sum_{i=1}^{N_A} p_i \, \mathbf{a}_i \odot \mathbf{a}_i} \tag{S27}$$

with respect to $\mathbf{u}$ of the objects $i \in A$ can be expressed in terms of the second order gyration moment as

$$R_\mathbf{u} = \sqrt{\mathrm{trace}\left(\mathbf{G}_{A,\mathbf{u}}^{(2)}\right)}. \tag{S28}$$

Just like the multipole moments[1] also the gyration moments (S26) can be easily shifted to another reference point, e.g. from the origin $\mathbf{0}$ to a local reference point $\mathbf{u}$. Choosing $\mathbf{u} = \mathbf{0}$ yields the gyration moments

$$\mathbf{G}_{A,\mathbf{0}}^{(n)} = \sum_{i=1}^{N_A} p_i \, \mathbf{r}_i^{(n)} \tag{S29}$$

with respect to the origin. Trivially, the zero'th order moments are independent of the reference point, i.e. $\mathbf{G}_{A,\mathbf{u}}^{(0)} = \mathbf{G}_{A,\mathbf{0}}^{(0)} = 1$. Inserting $\mathbf{a}_i = \mathbf{r}_i - \mathbf{u}$ into the definition (S26) of the local gyration moments one finds for the first moment

$$\mathbf{G}_{A,\mathbf{u}}^{(1)} = -\mathbf{u} + \sum_{i=1}^{N_A} p_i \, \mathbf{r}_i. \tag{S30}$$

According to the definition (S29), the second term is the first order gyration moment $\mathbf{G}_{A,\mathbf{0}}^{(1)}$ with respect to the origin and is identical to the center of geometry of cluster A as calculated in the global coordinate system, i.e. $\mathbf{G}_{A,\mathbf{u}}^{(1)} = -\mathbf{u} + \mathbf{G}_{A,\mathbf{0}}^{(1)}$. Analogously one finds for the second moment

$$\mathbf{G}_{A,\mathbf{u}}^{(2)} = \mathbf{u} \otimes \mathbf{u} - \mathbf{u} \otimes \mathbf{G}_{A,\mathbf{0}}^{(1)} - \mathbf{G}_{A,\mathbf{0}}^{(1)} \otimes \mathbf{u} + \mathbf{G}_{A,\mathbf{0}}^{(2)}. \tag{S31}$$

Thus, if one calculates gyration moments $\mathbf{G}^{(n)}_{A,\mathbf{0}}$ in a global coordinate system, then the local moments $\mathbf{G}^{(n)}_{A,\mathbf{u}}$ are easily obtained from the vector $\mathbf{u}$ marking the local reference point.

Now consider a second cluster B with the gyration moments $\mathbf{G}^{(n)}_{B,\mathbf{0}}$ and with $A \cap B = \emptyset$. The radius of gyration of the super-cluster $A \cup B$ with respect to the reference point $\mathbf{u}$ can now be conveniently computed from the second moments $\mathbf{G}^{(2)}_{A,\mathbf{u}}$ and $\mathbf{G}^{(2)}_{B,\mathbf{u}}$, obtained through Eq. (S31) as

$$R_{\mathbf{u}} = \sqrt{\text{trace}\left(W_A \mathbf{G}^{(2)}_{A,\mathbf{u}} + W_B \mathbf{G}^{(2)}_{B,\mathbf{u}}\right)/(W_A + W_B)}, \tag{S32}$$

where W_A and W_B are the norms defined by (S25). This argument is readily generalized to more than two clusters forming a super-cluster.

For charge distributions representing atomic clusters at the SAMM levels $l \geq 1$, which may contain also mixtures of PMM and DFT structural units, weights $w_a = 1$ are attached to all atoms a such that the norm W_A belonging to a cluster A is simply the number N_A of its atoms.

As compared to the analogous calculations[1] for constructing and shifting the multipole tensors $\mathbf{M}^{(n)}$, the computational effort for the gyration moments $\mathbf{G}^{(n)}$ is almost negligible, because the corresponding expressions are simpler and are only required up to second order, whereas the $\mathbf{M}^{(n)}$ are computed up to order $n = 4$.

S3. DFT/PMM ELECTROSTATICS: ENERGY CONSERVING FMM

It is a key feature of the $\text{SAMM}^{\chi,\text{E}}_{4,q,1}$ family[2] of FMM algorithms that the calculated forces do not only obey Newton's reaction principle,[1] but can also conserve the energy.[3] For grid-based DFT/PMM combinations a violation of energy conservation can be caused by so-called grid forces, which arise whenever a DFT atom moves relative to the DFT grid. These violations can be avoided, if the grid within the DFT box represents a cut-out part of an infinite and spatially fixed grid.[4] We adopt this procedure together with the most recent energy conserving algorithms[3] $\text{SAMM}^{\chi,\text{E}}_{4,q,1}$ for the computation of the electrostatic interactions between the DFT and PMM fragments.

As is explained in Section II C of the main text, in the DFT fragment the lowest cluster level $l = -3$ consists of small voxels λ, which disjointly partition the DFT grid and contain corresponding parts of the fragment's electron density. Utilizing a Voronoi tessellation, atomic sub-volumes V_μ of the grid are assigned to the DFT atoms μ by the unification of

all nearest neighbor voxels $\lambda(\mu)$. The unification of all atomic volumes V_μ, which belong to a structural unit[2,5] u, i.e. to a predefined cluster of atoms $\mu(u)$, then define the grid portion V_u assigned to this unit and so forth.

The algorithms employed for the FMM computation of the electrostatic interactions between DFT and PMM units or between larger atomic clusters have been previously described in great detail.[2–4] Thus it solely remains to be specified, how the electrostatic interactions of DFT voxels $\lambda(\mu)$ at levels $l = -3$, -2 and of DFT atoms μ at level $l = -1$ with PMM atoms are calculated.

1. Import of the PMM Potential Φ

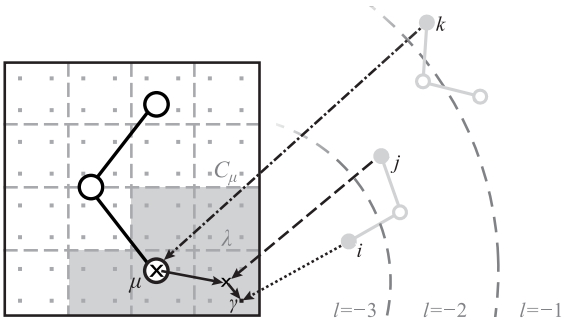


FIG. S16. Evaluation of Φ at a grid point γ. The construction of the level boundaries (dashed circle segments) is sketched in Section II C of the main text. See the text for further explanations.

Figure S16 explains how the external PMM potential Φ is computed at the highlighted grid point γ, which belongs to the voxel λ in the lower right corner of the DFT box. This voxel is part of the volume V_μ (gray shaded), which contains the charge distribution C_μ associated to the DFT atom μ (large cross). Three PMM atoms i, j, and k (large gray dots) illustrate increasingly distant sources of $\Phi(\mathbf{r}_\gamma)$.

The PMM atom i is too close to the center $\mathbf{r}_\lambda$ (small cross) of the voxel λ to pass the FMM voxel IAC (5). Therefore its contribution to $\Phi(\mathbf{r}_\gamma)$ is exactly calculated at level $l = -3$ from the potentials of its Gaussian charge and/or induced dipole distributions (dotted arrow).

The PMM atom j is sufficiently distant to pass this IAC [and additionally the Gaussian truncation IAC (6), which is not illustrated by Fig. S16 but by Fig. 2], but not distant enough from $\mathbf{r}_\mu$ to pass also the atomic IAC (7). Consequently, coefficients of a fourth order Taylor expansion at the center of the voxel λ are calculated at level $l = -2$ from the potentials

of the charge and/or induced dipole at the PMM atom j (dashed arrow), which are now treated as point objects. This expansion is evaluated at $\mathbf{r}_\gamma$ to approximate the contribution of j to $\Phi(\mathbf{r}_\gamma)$ (short solid arrow).

Because the most distant PMM atom k passes the atomic IAC (7), fourth order Taylor expansion coefficients are calculated at level $l = -1$ from its charge and induced dipole at $\mathbf{r}_\mu$ (dashed-dotted arrow). By a shifting procedure[1] (long solid arrow) these coefficients are transferred to the voxel center $\mathbf{r}_\lambda$, where they are added to the Taylor expansion coefficients generated by j. Hence the evaluation of the thus complemented Taylor expansion around $\mathbf{r}_\lambda$ at $\mathbf{r}_\gamma$ properly accounts also for the electrostatic action of the PMM atom k. In FMM the top-down inheriting of Taylor expansion coefficients from higher to lower levels is hierarchically repeated.[1,4]

2. Electrostatic Action of the DFT Electron Density

Once Φ has been calculated at all grid points and the KS orbitals are converged, the DFT electron density ρ_e is available through the grid charges q_γ. The associated Hellmann-Feynman forces on the charges and induced dipoles of the PMM atoms are calculated by operations, which represent exact inversions of the above import procedure.[4] Thus, these operations can be visualized by simply inverting all arrows in Fig. S16, which then acquire a different but complementary meaning.

The action of the charge q_γ located on the highlighted grid point γ in Fig. S16 on the PMM atom i is obtained by calculating at level $l = -3$ the potential (and its derivates) generated by it (inverted dotted arrow) at the atomic position $\mathbf{r}_i$ after shifting[4] the Gaussian shapes of the q_i and/or $\mathbf{p}_i$ to q_γ.

For interactions with the PMM atom j on level $l = -2$, a multipole expansion up to order $p = 4$ of all grid charges q_γ enclosed by the voxel λ is performed around $\mathbf{r}_\lambda$ (inverted solid arrow from $\mathbf{r}_\gamma$ to $\mathbf{r}_\lambda$). From these voxel multipole moments the potential (and derivatives) are computed at the PMM atom j (inverted dashed arrow), which render contributions of the $q_\gamma \in \lambda$ to the energy of and forces on atom j.

The multipole moments of all voxels $\lambda \in V_\mu$ are subsequently shifted (inverted long solid arrow) from the original reference points $\mathbf{r}_\lambda$ to the position $\mathbf{r}_\mu$ of the DFT atom μ, where they are summed up to yield the multipole moments of the atomic charge distribution C_μ.

Then, the potential (and derivatives), which are generated by these multipole moments, are calculated at the PMM atom k (inverted dashed-dotted arrow; level $l = -1$) to obtain the atomic energy and force. This procedure is hierarchically repeated at all higher levels $l \geq 0$.[1,4].

The apparent symmetry of Taylor and multipole expansions leads, in combination with the recent extensions,[3] to a Hamiltonian DFT/PMM electrostatics coupling scheme based on $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$ expansions,[1,4] as will be verified below.

S4. CONSERVATION OF ENERGY

Our DFT/PMM approach guarantees energy conservation as long as the association of atoms or clusters to hierarchy levels is unchanged, because it employs the most recent algorithms $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$, in which the forces are calculated as exact negative gradients of the energy.[3] Violations solely occur, whenever the dynamics causes a migration of atoms within the hierarchical tree of atomic clusters, because then the correspondingly altered approximation entails small temporal discontinuities of the calculated forces, which generate a certain amount of algorithmic noise.[2]

In earlier versions of SAMM[1,2] the forces obeyed Newton's third law but represented only approximate derivatives of the energy. Here, relative motions of clusters as measured by derivatives of the vector connecting their centers were neglected. This approximation entailed small violations of energy conservation. In a DFT/PMM-MD simulation, in which the cluster hierarchy and the connecting vector were kept fixed, these violations were shown to be absent.[4] In the following we employ a DFT/PMM simulation system, which features sizable relative motions of atomic clusters and concurrently an invariant association of atoms to FMM interaction levels. The combination of our revised interface with the most recent SAMM version[3] should conserve the energy for such a system.

A. Methods

Figure S17 depicts a DFT/PMM sample simulation system specifically designed to check the above claim. It comprises one alanine dipeptide molecule making up the DFT fragment and three polarizable TL6P water molecules[6] representing the PMM fragment. The motions

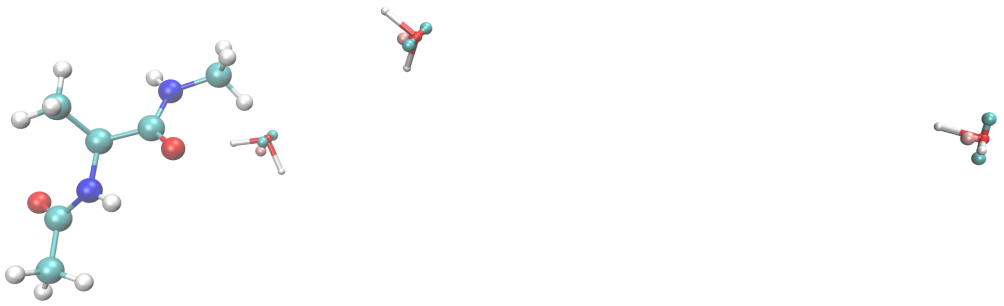


FIG. S17. DFT/PMM simulation system for a check of energy conservation in which the alanine-dipeptide molecule represented the DFT fragment and three TL6P water molecules[6] the PMM fragment.

of the molecules were constrained by spatially fixed harmonic potentials (of spring constant 1.0 kcal/mol Å^2), which were attached to the C_α atom of the alanine dipeptide and to the O atoms of the TL6P water molecules. The intermolecular distances were chosen such that interactions between the DFT fragment and the various PMM atoms were always calculated at the same FMM levels $l = -3, \ldots, 0$, respectively, during the whole 250 fs simulation. As a result, all associated algorithmic features were probed. The average temperature was 414 K, and the convergence criteria were tightened to $\chi_{\mathrm{DFT}} = 10^{-8}$ and $\chi_{\mathrm{PMM}} = 10^{-5}$ D. The remaining aspects of the simulations are described in Section III A.

Energy conservation was checked using the so-called shadow Hamiltonian technique,[7] which derives from the fact that a numerical integration of a Hamiltonian dynamics with the Verlet algorithm[8] does not conserve the energy E but a different temporal average quantity $\langle E(t\,|\,\Delta t)\rangle_\tau$, which depends on the integration time step Δt and on the width τ of the employed time window. In the limits $\Delta t \to 0$ and $\tau \to \infty$ one has $[\langle E(t\,|\,\Delta t)\rangle_\tau - E] \sim (\Delta t)^2 \to 0$.

For a MD simulation yielding the energy trajectory $E(t\,|\,\Delta t)$ this modified conservation law can be verified by checking whether the following equation[7]

$$E(t\,|\,\Delta t) = \langle E(t\,|\,\Delta t)\rangle_\tau + f(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{S33}$$

holds with a shape function $f(t)$, which is independent of Δt. Then the scaled energy fluctuations

$$\Delta E(t\,|\,\Delta t)/\Delta t^2 \equiv \left[E(t\,|\,\Delta t) - \langle E(t\,|\,\Delta t)\rangle_\tau\right]/\Delta t^2$$

are given by

$$\Delta E(t\,|\,\Delta t)/\Delta t^2 = f(t) + \mathcal{O}(\Delta t^2) \tag{S34}$$

with the standard deviation

$$\sigma_{\Delta E} = \sqrt{\langle f^2(t)\rangle_\tau}\Delta t^2 + \mathcal{O}(\Delta t^4), \tag{S35}$$

which vanishes in the analytical limit $\Delta t \to 0$.

If our implementation is correct and, hence, yields energy conservation within the above simulation setup, then the DFT/PMM energy trajectories obey Eqs. (S34) and (S35). If two simulations use different time steps (e.g. Δt_0 and $\Delta t_0/2$) and otherwise start at identical conditions, then by Eq. (S34) the calculated shape functions $f(t)$ should be identical and by Eq. (S35) the standard deviation $\sigma_{\Delta E}$ should decrease quadratically with Δt.

B. Results

Figures S18a and S18b show the scaled energy fluctuations $\Delta E(t\,|\,\Delta t)/(\Delta t/\Delta t_0)^2$ observed in the sample simulations for the time steps $\Delta t_0 = 0.5$ fs and $\Delta t_0/2$ with the previous uncorrected $\mathrm{SAMM}^{\mathrm{f,O}}_{4,1,1}$ and the energy-gradient-corrected $\mathrm{SAMM}^{\mathrm{f,E}}_{4,1,1}$ algorithm, respectively. The fluctuations observed for $\Delta t = \Delta t_0$ are drawn by a red solid line, the scaled fluctuations $4\Delta E(t\,|\,\Delta t)$ observed for $\Delta t = \Delta t_0/2$ by a green dotted line.

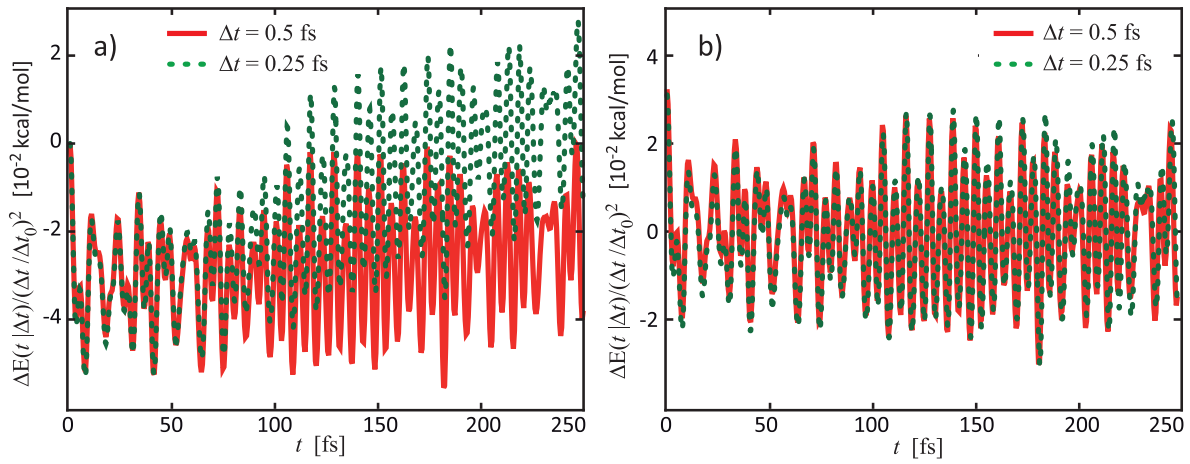


FIG. S18. Fluctuations of the total DFT/PMM energy during a 250 fs MD simulation of alanine-dipeptide (DFT fragment) surrounded by three PMM water molecules for different time steps $\Delta t = 0.5$ fs (red) and $\Delta t = 0.25$ fs (dotted green): a) previous uncorrected $\mathrm{SAMM}^{\mathrm{f,O}}_{4,1,1}$ version,[2] b) energy-gradient-corrected[3] $\mathrm{SAMM}^{\mathrm{f,E}}_{4,1,1}$.

After scaling the red and green fluctuations shown in Fig. S18a quickly deviate from each other indicating the violation of energy conservation by $\mathrm{SAMM}^{\mathrm{f,O}}_{4,1,1}$. In contrast, the fluctuations depicted by Fig. S18b exhibit the expected almost perfect match. Hence, for the given sample system the energy is conserved by the energy-gradient-corrected $\mathrm{SAMM}^{\mathrm{f,E}}_{4,1,1}$. The energy-corrected versions $\mathrm{SAMM}^{\chi,\mathrm{E}}_{4,q,1}$, therefore, have become our algorithms of choice for DFT/PMM simulations with the most accurate variant $\mathrm{SAMM}^{\mathrm{a,E}}_{4,3,1}$ representing our new default. Furthermore the comparison additionally demonstrates that the relative motions of cluster centers is the only cause for the violation of energy conservation with the previous SAMM_4 based DFT/PMM interface, which is apparent in Fig. S18a.

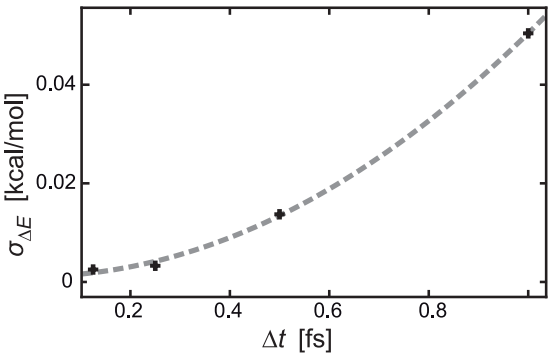


FIG. S19. Standard deviations $\sigma_{\Delta E}$ of the energy fluctuations $\Delta E(t\,|\,\Delta t)$ for different time steps $\Delta t \in \{0.125, 0.250, 0.500, 1.000\}$. The gray dashed lines shows a fitted parabola.

According to Eq. (S35) the standard deviation $\sigma_{\Delta E}$ of the energy fluctuations $\Delta E(t\,|\,\Delta t)$ should vanish quadratically with the time step size Δt. Figure S19 shows that this is actually the case already for the least accurate variant $\mathrm{SAMM}^{\mathrm{f,E}}_{4,1,1}$ as demonstrated by MD simulations with the time steps $\Delta t \in \{0.125, 0.25, 0.50, 1.00\}$ of alanine-dipeptide (DFT fragment) surrounded by three PMM water molecules. Small deviations from the fitted parabola are most likely due to the limited statistics of the sample trajectories (250 fs).

S5. THE SAMM$^{\chi,\mathrm{E}}_{4,1,1}$ ERROR $D_{\mathbf{FMM}}$ IS INDEPENDENT OF R_Λ

Section IV C of the main text starts with the claim that the accuracy, at which the revised DFT/PMM interface computes the grid potential $\Phi(\mathbf{r}_\gamma)$, does not depend on the voxel size R_Λ. For a check of this claim consider Figure S20, which demonstrates that the FMM errors $D_{\mathrm{FMM}}(\Theta_\chi)$ [cf. Eq. (14)] of the three SAMM$^{\chi,\mathrm{E}}_{4,1,1}$ variants are actually independent of the voxel size R_Λ.

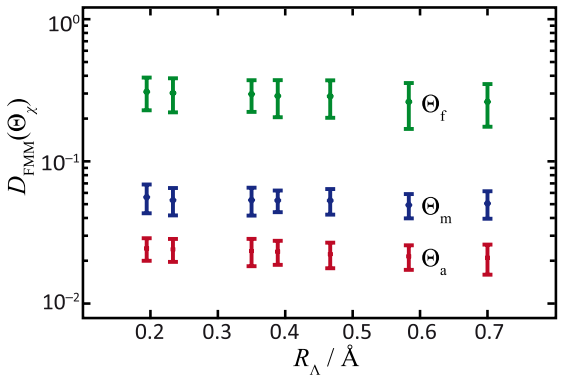


FIG. S20. Errors $D_{\mathrm{FMM}}(\Theta_\chi)$ defined by Eq. (14) of the SAMM$^{\chi,\mathrm{E}}_{4,1,1}$ algorithms ($\chi \in \{\mathrm{f}, \mathrm{m}, \mathrm{a}\}$, green, blue, red) are independent of the voxel size R_Λ.

S6. RECTANGULAR VOXELS

The main text is restricted to the discussion of cubic DFT boxes and cubic voxels. For non-cubic but still rectangular DFT boxes, also the voxels, which commensurably partition the DFT grid, are rectangular (cf. Sec. II C 1).

To estimate the sizes of the rectangular voxels, we disregard the enclosed charge densities, just like we did already in the cubic case. Thus, we choose all voxels of identical shape and size as prototypically represented by a rectangular volume element $V = l_x l_y l_z$ centered at the origin. Here, the side lengths l_α of the voxels are chosen such that they are as close as possible to $R_\Lambda^{\mathrm{opt}} \approx 0.35$ Å and that they are integer divisors of the associated length L_α of the DFT box. The geometric radius of gyration $R_V = \sqrt{\langle \mathbf{r}^2 \rangle_V}$ of a voxel is the standard deviation of the position vector $\mathbf{r}$ within V, where $\langle \ldots \rangle_V$ denotes the average over V. Note that Eq. (S27) is, with $p_i = 1$ and $\mathbf{u} = \mathbf{0}$, a discretized version of this expression. For a

rectangular volume element one gets

$$R = \frac{1}{2}\sqrt{(l_x^2 + l_y^2 + l_z^2)/3}, \tag{S36}$$

which reduces for a cubic voxel with $l_x = l_y = l_z = l$ to $R = l/2$.

REFERENCES

[1]K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, and P. Tavan, J. Chem. Theory Comput. **8**, 3628 (2012).

[2]K. Lorenzen, C. Wichmann, and P. Tavan, J. Chem. Theory Comput. **10**, 3244 (2014).

[3]K. Lorenzen and P. Tavan, (2015, in preparation).

[4]M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan, and G. Mathias, J. Chem. Phys. **138**, 244103 (2013).

[5]C. Niedermeier and P. Tavan, J. Chem. Phys. **101**, 734 (1994).

[6]P. Tröster, K. Lorenzen, and P. Tavan, J. Phys. Chem. B **118**, 1589 (2014).

[7]M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, 1st ed. (Oxford University Press USA, New York, 2010) Chap. 3.13, pp. 121–124.

[8]W. C. Swope and H. C. Anderson, J. Chem. Phys. **76**, 637 (1982).

2.3 Beschreibung der IR-Spektren von Amidgruppen mit DFT/PMM-MD

Die nachfolgende abgedruckte Publikation

A Polarizable QM/MM Approach to the Molecular Dynamics of Amide Groups Solvated in Water

Magnus Schwörer, Christoph Wichmann und Paul Tavan
J. Chem. Phys. **144**, 114504 (2016),

die ich zusammen mit Christoph Wichmann und Paul Tavan verfasst habe, beschreibt die Entwicklung und Evaluierung des GP6P-Wassermodells, die Optimierung der Lennard-Jones-Wechselwirkungen zwischen DFT- und PMM-Fragment, und die Berechnung der IR-Spektren von isolierten bzw. von in GP6P gelösten NMA-Molekülen aus ausgedehnten DFT-MD bzw. DFT/PMM-MD Trajektorien.

THE JOURNAL OF CHEMICAL PHYSICS **144**, 114504 (2016)

A polarizable QM/MM approach to the molecular dynamics of amide groups solvated in water

Magnus Schwörer,[a] Christoph Wichmann,[a] and Paul Tavan[b]
Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

(Received 29 January 2016; accepted 26 February 2016; published online 21 March 2016)

The infrared (IR) spectra of polypeptides are dominated by the so-called amide bands. Because they originate from the strongly polar and polarizable amide groups (AGs) making up the backbone, their spectral positions sensitively depend on the local electric fields. Aiming at accurate computations of these IR spectra by molecular dynamics (MD) simulations, which derive atomic forces from a hybrid quantum and molecular mechanics (QM/MM) Hamiltonian, here we consider the effects of solvation in bulk liquid water on the amide bands of the AG model compound *N*-methyl-acetamide (NMA). As QM approach to NMA we choose grid-based density functional theory (DFT). For the surrounding MM water, we develop, largely based on computations, a polarizable molecular mechanics (PMM) model potential called GP6P, which features six Gaussian electrostatic sources (one induced dipole, five static partial charge distributions) and, therefore, avoids spurious distortions of the DFT electron density in hybrid DFT/PMM simulations. Bulk liquid GP6P is shown to have favorable properties at the thermodynamic conditions of the parameterization and beyond. Lennard-Jones (LJ) parameters of the DFT fragment NMA are optimized by comparing radial distribution functions in the surrounding GP6P liquid with reference data obtained from a "first-principles" DFT-MD simulation. Finally, IR spectra of NMA in GP6P water are calculated from extended DFT/PMM-MD trajectories, in which the NMA is treated by three different DFT functionals (BP, BLYP, B3LYP). Method-specific frequency scaling factors are derived from DFT-MD simulations of isolated NMA. The DFT/PMM-MD simulations with GP6P and with the optimized LJ parameters then excellently predict the effects of aqueous solvation and deuteration observed in the IR spectra of NMA. As a result, the methods required to accurately compute such spectra by DFT/PMM-MD also for larger peptides in aqueous solution are now at hand. © *2016 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4943972]

I. INTRODUCTION

Hybrid approaches combining a quantum mechanical (QM) description of a solute molecule with a molecular mechanics (MM) force field for its solvent environment[1] have become a popular and powerful tool for the theoretical investigation of (bio)molecular properties.[2–4] As compared to pure QM molecular dynamics (MD) simulations[5–7] such hybrid QM/MM techniques allow to extend the sizes of the simulated systems and the accuracy of the statistical sampling by orders of magnitude.

Building upon and substantially revising earlier methods[8,9] a Hamiltonian hybrid approach for MD simulations has been recently presented,[10,11] which combines a density functional theory[12,13] (DFT) treatment of a solute molecule with a polarizable molecular mechanics (PMM) force field for its solvent environment by coupling the parallelized grid-based plane-wave DFT code CPMD[14] to the parallelized PMM-MD program IPHIGENIE.[15] This method computes the interaction forces between the DFT and PMM fragments efficiently from the DFT/PMM interaction Hamiltonian by employing an energy conserving fast multipole method called the structure adapted multipole method (SAMM).[16–18] Its present implementation enables stable hybrid MD simulations on multi-core high-performance computing systems.

Just like its early DFT/MM predecessor,[8] which had been restricted to common non-polarizable MM force fields, also the new DFT/PMM approach has been particularly designed to calculate the infrared (IR) spectra of biomolecules in condensed phase environments (for reviews over previous DFT/MM studies addressing such IR spectra see Refs. 19–21). In this contribution, we will focus on the IR spectrum of *N*-methyl-acetamide (NMA) in aqueous solution (the chemical structure of NMA is depicted in Figure 1), which has been extensively studied by QM/MM methods embedding NMA either in the bulk liquid[22–24] or in water clusters[25] and by pure QM approaches toward NMA in the liquid[6,26,27] and in clusters[28–31] (for a recent review of available literature see Ref. 31). NMA in water is important, because it represents a minimal model for the amide groups (AGs) making up the backbones of the polypeptides and because the native environment of these biopolymers predominantly consists of water. AGs are highly polar and polarizable. Correspondingly, the spectral locations of the IR amide bands, which arise from vibrational transitions within the AGs, are steered by the polarizing electric fields generated by their respective condensed phase environments.[24,29,32]

[a]M. Schwörer and C. Wichmann contributed equally to this work.
[b]Electronic mail: tavan@physik.uni-muenchen.de

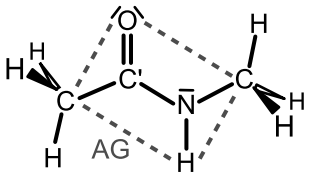


FIG. 1. Chemical structure of NMA. Gray dashed lines indicate the planar amide group (AG).

The electric forces acting between NMA and the surrounding water mainly arise from hydrogen bonding interactions of the water with the AG core of NMA (cf. Figure 1), which has one acceptor (C′═O group) and one donor (N—H group) of hydrogen bonds. The strengths of these hydrogen bonding interactions shape the distribution of the water molecules in the vicinity of the AG and, thus, the strength of the polarizing electric reaction field. They depend on both the electrostatic and the van der Waals interactions between the atoms C′, O, N, and H of the AG and the surrounding water molecules.

For a DFT description of NMA in an aqueous PMM environment, one has to choose a DFT setup for the NMA, a PMM force field for the water molecules, and, as is common in QM/MM methods,[2] two-parameter Lennard-Jones (LJ) potentials $U_{\mathrm{LJ}}(r \mid A, B) = A/r^{12} - B/r^{6}$, which model the van der Waals interactions between the DFT and PMM fragments. If, in the chosen PMM force field, all sources of the electrostatic potential are represented by Gaussian distributions (see further below), then solely the LJ parameters of the four AG atoms remain for a tuning of the hydrogen bonding strengths. In contrast, the LJ parameters of the remaining atoms of NMA, e.g., in the methyl groups, should have only a minor effect on these strengths. They can be safely adopted from a standard MM force field like CHARMM22.[33]

Here we will address the important question as to how one can optimize the LJ parameters of the four AG atoms for an adequate modeling of the hydrogen bonding strengths between AGs and an aqueous PMM environment. Such a question arises in all QM/MM applications,[34–38] because the van der Waals interactions between the MM and QM fragments are generally described by empirical energy functions,[1,2] whose parameters have to be separately optimized for a physically realistic description of the forces acting between the MM and QM fragments.

The intended optimization of the enumerated LJ potentials requires a reference observable for the hydrogen bonding strengths. We will choose the radial distribution functions (RDFs) $g_{\mathrm{OH_w}}(r)$ of the water hydrogens $\mathrm{H_w}$ around the O atom and $g_{\mathrm{HO_w}}(r)$ of the water oxygens $\mathrm{O_w}$ around the H atom of the AG, respectively, at inter-atomic distances r. Because corresponding experimental data are unknown to us, we will compute reference RDFs by an empirically dispersion-corrected[39] "first-principles" DFT-MD simulation,[7,40] which should yield reasonably reliable information about their first peaks. It will then be interesting to see to what extent the optimization of the AG's LJ parameters can bring the DFT/PMM-MD results close to the reference RDFs and how it affects the DFT/PMM predictions on the IR spectra of NMA in water.

It is well known that suitably scaled DFT treatments of isolated organic molecules can accurately describe their gas phase IR spectra.[41,42] Therefore one may expect that also IR spectra calculated for NMA in water from DFT/PMM-MD simulations can match corresponding spectroscopic data[29] quite well, if the RDFs from DFT/PMM-MD are close to the references and if the employed PMM model of water describes the bulk phase properties of liquid water with a reasonable accuracy.

The use of non-polarizable MM water models such as TIP3P or TIP4P,[43] which provide poor descriptions[44] of the bulk liquid properties, can diminish the quality of IR spectra computed by DFT/MM techniques.[7,24,45] The reason is that the IR spectra of strongly polar and polarizable molecules in likewise polar and polarizable condensed phase environments represent extremely sensitive probes for the details of the interactions between the molecule and its environment.[21] Here, for instance, the mutual polarization of solute and solvent has been demonstrated to have large effects on the IR spectra calculated by DFT/MM hybrid methods.[46,47] In this respect, the new DFT/PMM technology[10,11] sketched above represents an important progress, because it does not only account for the electronic polarization of the DFT fragment but also for that of the PMM solvent.

For a high-quality match of DFT/PMM predictions on the IR spectra of polar and polarizable molecules in liquid water with corresponding spectroscopic data, one should thus choose a PMM force field, which models the physical properties of the bulk liquid quite well and, concurrently, is specifically adapted to DFT/PMM settings. Technically, the new DFT/PMM approach has enabled the computational derivation and efficient handling of relatively complex PMM water models featuring many points of force action,[48,49] which properly account[50] for important properties of the bulk liquid. This is fortunate for our present purpose of describing the IR spectrum of NMA in water, because we now can apply this technology for the construction of a PMM water model potential specifically designed for DFT/PMM hybrid descriptions.

Whenever one applies a grid-based DFT approach, as implemented, e.g., in CPMD,[14] for hybrid calculations one should avoid force fields, which model the electrostatic signatures of the molecular structures in the (P)MM fragment by point charges or point dipoles. The reason is that condensed phase biomolecular systems always feature (P)MM atoms inside the DFT box harboring the grid-discretized electron density of the DFT fragment. If these (P)MM atoms carry point sources of the electrostatic potential, they can spuriously distort this electron density.[51–53] This unphysical distortion can be avoided by artificially smoothing the potentials of these point sources, whenever one computes interactions between the DFT and (P)MM fragments.[8–10,54,55] Such a smoothing generally introduces additional scale parameters, which strongly modify the near-field interactions between the two fragments, as one can observe, for instance, by their effects on RDFs of (P)MM solvent molecules surrounding a DFT solute.[9]

The generic way for effecting a smoothing on a length scale σ is the replacement of an electrostatic point source by

a Gaussian distribution of width σ. Therefore, DFT/(P)MM calculations should preferentially employ a (P)MM force field, in which all charges and dipoles are represented by Gaussian distributions.

At least two such PMM water models have been previously suggested, a three-point so-called "charge-on-spring" model[56] and an earlier four-point model exhibiting three static Gaussian partial charges and an induced Gaussian dipole.[57] Both feature planar charge distributions. However, an adequate modeling of the water molecule's quadrupole moment, which shapes the short-range interactions and, hence, the local order structures in the liquid phase, requires additional static partial charges, the so-called "lone-pair" charges, symmetrically situated above and below the molecular plane.[58] Correspondingly we will apply the DFT/PMM technique to parameterize a "Gaussian Polarizable 6-Point" (GP6P) model exhibiting five static Gaussian partial charges and one Gaussian induced dipole.

Note here that the difference between the electrostatic potentials of point and Gaussian sources becomes numerically irrelevant (single precision) at distances $r/\sigma \geq 6$. For typical smoothing scales of $\sigma \lesssim 0.8$ Å this distance becomes as small as 4.8 Å. If one exploits the numerical equivalence between the two types of potentials at larger distances, one should be able to keep the additional cost of computing potentials of Gaussian sources quite limited. Moreover, by applying the linearly scaling fast multipole method SAMM[16–18] to the approximate computation of the long-range electrostatic and van der Waals interactions, also locally very complex PMM water models can be efficiently treated,[59] because the costly exact evaluation of pair interactions can be replaced by fast multipole descriptions already at distances as small as 5.4 Å. Therefore we expect that the corresponding MD program IPHIGENIE[15] will be capable to treat also the new GP6P model with a reasonable efficiency despite the sizable complexity of GP6P.

This contribution is organized as follows. First we will shortly sketch some of the new aspects entering the construction of the GP6P water model. Because key concepts of the employed parameterization strategy have been explained previously,[48,49] the detailed description of the applied methods has been moved to Section S1 of the supplementary material.[60]

Subsequently we describe the first-principles DFT-MD simulations of NMA in a small periodic box additionally containing 64 water molecules, which served us to estimate the shape of the RDFs $g_{OH_w}(r)$ and $g_{HO_w}(r)$ for hydrogen-bonding distances $r \in [1.5, 2.5]$ Å. The corresponding results will then provide reference data for a computational optimization of the LJ parameters associated with the AG atoms in the DFT model of NMA. For reasons of computational manageability, we will apply a DFT/PMM mean-field approach in this optimization. Then we will describe the DFT/PMM-MD simulations, from which we compute the IR spectra of NMA in liquid water by Fourier-transform techniques.[20]

As our first result, we sketch the structure and liquid phase properties of the new GP6P water model. Next we show how the RDFs obtained from DFT/PMM-MD change upon the optimization of the LJ potentials and how these changes are reflected in the IR spectra of solvated NMA derived from DFT/PMM-MD. Using the optimized LJ parameters, we finally compare IR spectra from DFT/PMM-MD with corresponding experimental spectra and with results of previous DFT/MM-MD simulations.[24] For these comparisons, method-specific scaling factors are required, which we derive from DFT-MD simulations of NMA isolated in the vacuum for three different DFT functionals.

II. METHODS

The (P)MM- and DFT/PMM-MD simulations of NMA in periodic boxes filled with (P)MM water were carried out with the hybrid program IPHIGENIE/CPMD, which integrates the parallel grid-based plane-wave DFT code CPMD[14] into the parallel PMM-MD code IPHIGENIE,[10,11,15–18,61] whereas the first-principles DFT-MD simulation of NMA solvated in a much smaller box of liquid water was performed with the CP2K software package.[62,63]

A. General MM-, PMM-, and DFT/PMM-MD settings

In all (P)MM- and DFT/PMM-MD simulations the geometries of the water molecules were constrained using the M-SHAKE[64] and RATTLE[65] algorithms with relative tolerances of 10^{-10}. The equations of motion were integrated with the velocity Verlet algorithm[66] employing a time step of 1 fs in the (P)MM or of 0.5 fs in the DFT/PMM settings, respectively. In these MD simulations the pressure was calculated from the virial expression.[67,68]

Long-range forces were treated by the most recent energy-conserving version of the SAMM[16–18] employing 4th and 3rd order symmetric Taylor expansions for the electrostatic and van der Waals dispersion interactions, respectively. The transition from the exact evaluation of the associated pair expressions to the approximate SAMM descriptions is steered by the parameter Θ entering the SAMM interaction acceptance criterion,[17] for which we chose $\Theta_m = 0.20$ (intermediate accuracy) in the (P)MM settings and $\Theta_a = 0.17$ (high accuracy) in the DFT/PMM settings.

The SAMM expansions were applied up to a maximum distance d_{MIC} dictated by the minimum image convention (MIC) of the employed toroidal boundary conditions.[69] To cover the electrostatics also at distances beyond d_{MIC}, a moving-boundary reaction field approach[70] was used, which describes a surrounding dielectric continuum with a dielectric constant of 78. For the corresponding long-range parts of the dispersion attraction a continuum correction was applied.[71] In the PMM and DFT/PMM settings the self-consistency of the induced dipoles was defined by the convergence threshold[10] $\chi_{PMM} = 5 \times 10^{-5}$ D.

B. Parameterization and evaluation of GP6P

Employing these general simulation settings the new DFT/PMM hybrid technology was applied to the parameterization of the new GP6P model for water. Due to most recent advances[11,17,18] this technology has meanwhile come to maturity, such that the associated PMM-MD program

package IPHIGENIE[15] is now available to the public (the current distribution of the DFT program CPMD[14] contains additionally required code).

This availability is important, because it guarantees that all simulations presented in this work have become reproducible and extendable for the interested scientific community. Moreover, particularly concerning the GP6P water model, for which the download[15] contains sample simulation systems, the computations can be extended to the study of further properties of the bulk liquid (or the solid and gas phases) beyond the limited set of observables investigated here for a first quality assessment.

Because most procedures and physical concepts applied to the parameterization of GP6P are adopted from previous work,[48,49,72,73] these issues have been transferred to Section S1 of the supplementary material.[60] Driven by the aim to support future parameterizations of complex PMM models for other liquids, this section presents a thorough and detailed description of the applied methods and, in passing, corrects previous[48,49] ambiguities and errors.

The parameters of the GP6P model were obtained by three different approaches, i.e., by

(i) the direct adoption of well-established experimental values (molecular geometry in the liquid phase,[74,75] gas phase dipole moment,[76] and polarizability,[77] cf. Section S1 A 1[60]),
(ii) DFT/PMM-based derivations of almost all electrostatic parameters (Gaussian width of the induced dipole distribution; magnitudes, Gaussian widths, and positions of the static partial charge distributions, cf. Sections S1 C 1-S1 C 5[60]), and
(iii) PMM-MD-based empirical optimization[78] of one Gaussian charge width and of the three parameters A_1, A_2, and B of the Buckingham potential[79] [Eq. (S1)[60]], which models the GP6P van der Waals interactions (cf. Sections S1 C 6 and S1 C 7[60]).

The PMM-MD simulations were conducted in a periodic box containing $N_m \equiv 1500$ GP6P models (cf. Section S1 B 1[60]) at the experimental density[80] $n^{exp}(p_0,T_0) = 0.9965$ g/cm^3 assumed by liquid water at the temperature $T_0 \equiv 300$ K and the pressure $p_0 \equiv 1$ atm. As experimental target values for the empirical optimization of the four parameters enumerated in (iii) we chose the mean potential energy $E_0^{exp} = -9.92$ kcal/mol per molecule,[81] the position[82] $r_{OO}^{max,exp} = 2.76$ Å of the first peak of the oxygen-oxygen RDF $g_{OO}(r)$, the pressure p_0, and the isobaric thermal expansion coefficient[80] $\alpha_p^{exp} = 2.8 \times 10^{-4}$ K^{-1}.

Note that Section S1 D of the supplementary material[60] contains, mainly through references to previous work, a sketch of the methods applied in the computation of several GP6P predictions on important liquid phase properties.

C. Simulation systems for NMA in water

All MD simulations of NMA in water employed cubic periodic simulation boxes. For MM/(P)MM-MD simulations large boxes were filled with 4494 TIP3P[43] or GP6P water models and one CHARMM22 model[33] of NMA. These MM/(P)MM simulation systems were equilibrated by 1 ns MD simulations in the Np_0T_0 ensemble employing a Berendsen barostat[83] (coupling time 10 ps) and a Bussi thermostat[84] (coupling time 0.1 ps), respectively. The resulting simulation box volumes were $V_{MM} = (51.5\ \text{Å})^3$ and $V_{PMM} = (51.3\ \text{Å})^3$ corresponding to the densities of 0.983 g/cm^3 and 0.998 g/cm^3, respectively.

For the first-principles DFT-MD simulation a small box containing — beyond NMA — also 64 water molecules was constructed from a subsequent 1 ns MD simulation of the MM simulation system (CHARMM22/TIP3P) in the $NV_{MM}T_0$ ensemble. In 1000 snapshots of this $NV_{MM}T_0$ ensemble a small cubic box of varying volume was centered around the NMA in such a way that it contained 64 water oxygen atoms.[7] The resulting average volume $V_{DFT} = (12.7\ \text{Å})^3$, which corresponds to a density of 0.994 g/cm^3, was then chosen for the DFT-MD simulation. From one of the snapshots the CHARMM22-NMA and 64 surrounding TIP3P water models were taken after MM minimization of the associated cluster as initial conditions for DFT-MD.

To obtain the initial conditions for DFT/PMM-MD simulations, we performed a 1 ns MD simulation of the MM/PMM system (CHARMM22/GP6P) in the $NV_{PMM}T_0$ ensemble and drew 24 phase-space snapshots from the tail of the trajectory at temporal distances of 5 ps, which we collected into the set $\mathcal{S}_{PMM}$.

D. First-principles MD simulation

For our 100 ps first-principles Born-Oppenheimer MD simulation of NMA surrounded by 64 water molecules we adopted the methods described in Ref. 7. Thus, the MD simulation was performed with the CP2K program package[62,63] using the exchange functional of Becke,[85] the correlation functional of Lee, Yang and Parr[86] (BLYP) and the norm-conserving pseudopotentials proposed by Goedecker, Teter, and Hutter.[87–89]

The Kohn-Sham orbitals were expanded in a triple-ζ valence basis set including double polarization functions.[62] Electrostatic interactions were treated using the Gaussian and plane waves scheme[90] implemented in the Quickstep[62] module of CP2K[63] at a density cutoff of 400 Ry. The self-consistent field procedure employed the orbital transformation method suggested in Ref. 91 and a convergence criterion of 10^{-7}.

In contrast to a similar calculation on solvated NMA performed by Gaigeot *et al.*,[6] we used an empirical dispersion correction scheme. Here, we adopted the suggestion by Grimme,[39] which introduces an inter-atomic attraction $\sim r^{-6}$. This DFT-MD simulation model has been shown to yield a reasonable structure and the proper density of liquid water.[92,93]

Taking the configuration obtained from MM-MD (cf. Section II C) as the starting structure the system was simulated for 100 ps (of which the first 15 ps were excluded from the subsequent computation of RDFs) in the $NV_{DFT}T_0$ ensemble with an integration time step of 0.5 fs. Here the temperature T was controlled by a massive Nose-Hoover chain thermostat (chain length 3, time constant of 1 ps).[94] The average pressure in this small fixed-volume system turned out to be 44 ± 250 atm. With the help of the program VMD[95] the desired

reference RDFs $g^{\mathrm{ref}}_{\mathrm{OH_w}}(r)$ and $g^{\mathrm{ref}}_{\mathrm{HO_w}}(r)$ of the water molecules around the C′═O and the N—H groups of NMA, respectively, were eventually calculated in the range [1.4 Å, 3.0 Å] with a bin size of 0.05 Å from the last 85 ps of the trajectory.

E. DFT-MD and DFT/PMM-MD simulations of NMA

With the aim to optimize the LJ parameters of AG atoms in the DFT fragment NMA we performed DFT/PMM simulations in the $NV_{\mathrm{PMM}}T_0$ ensemble. The quality of the thus established description was checked by computing the IR spectrum of NMA in aqueous solution from extended DFT/PMM-MD simulations. Because comparisons of DFT-based predictions with experimentally observed vibrational frequencies require a method-specific scaling factor f_{DFT}, we calculated the gas phase IR spectrum of NMA from DFT-MD simulations at T_0 *in vacuo* and compared the results with gas phase IR data.[29] These simulations cover comparable amounts of anharmonic frequency shifts as DFT/PMM-MD simulations of NMA in GP6P water.

1. DFT setups for NMA in vacuo and in PMM water

We employed three different DFT setups for the description of the NMA molecule, namely, (i) the BLYP,[85,86] (ii) B3LYP[86,96–98] functionals combined with norm-conserving Martins-Troullier[99] (MT) pseudopotentials and a plane-wave cutoff at 70 Ry, and (iii) the exchange functional of Becke[85] in connection with the correlation functional of Perdew[100] (BP), the MT pseudopotentials, and a 80 Ry cutoff. We denote these DFT setups as MT/BLYP, MT/B3LYP, and MT/BP, respectively. The latter setup was selected to enable comparisons with the IR spectra of NMA in TIP4P water[43] calculated by Schultheis *et al.*[24] from DFT/MM simulations, which also applied MT/BP to NMA.

For all three functionals, the MT pseudopotentials model the core-valence interaction by *s* and *p* potentials for C, O, and N atoms with the respective radii for *s* and *p* chosen equal as 1.23 a.u., 1.05 a.u., and 1.12 a.u., while H atoms are treated as a single *s* potential with a radius of 0.5 a.u. The MT pseudopotentials for B3LYP were chosen identical to those for BLYP. In all DFT- and DFT/PMM-MD simulations the NMA molecule was centered in a cubic DFT box with a volume of $(11\ \text{Å})^3$. The DFT self-consistency convergence criterion[10] was $\chi_{\mathrm{DFT}} = 10^{-6}$.

2. DFT dynamics simulations of isolated NMA

For the isolated NMA molecule we first conducted 12 Langevin dynamics simulations for each of the three DFT functionals spanning 10 ps each. These simulations employed a second-order integration algorithm[101] controlling the target temperature T_0 with a decay time of 1 ps. The 12 initial structures of NMA were adopted from the set $\mathcal{S}_{\mathrm{PMM}}$ (cf. Section II C). The resulting phase space structures of isolated NMA were subsequently taken for each of the three functionals as the initial conditions for 12 MD simulations, each of 50 ps duration, in which translations and rotations of the molecule were suppressed. The rotation correction is necessary, because the DFT grid of CPMD destroys the isotropy of the simulated system and, therefore, prevents the conservation of the total angular momentum. From these trajectories we subsequently calculated gas phase IR spectra of NMA by a Fourier transform technique (see Section II F 3).

For the MT/BLYP setup additional simulations of this kind (12×10 ps Langevin, 12×50 ps MD) were carried out for the target temperatures 20 K, 75 K, 150 K, and 433 K. The gas phase IR spectra derived from these simulations served to estimate the sizes of the anharmonic frequency shifts at T_0.

3. DFT/PMM-MD of NMA in water

All 24 members of the snapshot set $\mathcal{S}_{\mathrm{PMM}}$ (see Section II C) were equilibrated for 10 ps in the $NV_{\mathrm{PMM}}T_0$ ensemble for each of the three DFT/PMM hybrid setups, in which NMA was described either by MT/BLYP, MT/BP, or by MT/B3LYP. Here, T_0 was tightly controlled (coupling time 0.1 ps) by a Bussi thermostat,[84] which was, like in all DFT/PMM-MD simulations, exclusively coupled to the solvent degrees of freedom.[102] The end points of these equilibrations served as initial conditions for 50 ps DFT/PMM-MD simulations in the $NV_{\mathrm{PMM}}T_0$ ensemble. Here, the thermostat coupling time was increased to 1 ps.

In all three DFT/PMM descriptions the LJ potentials of AG atoms in NMA were specified by an optimized parameter set $\mathcal{L}_{\mathrm{opt}}$ (cf. Section II F). Employing the LJ parameters $\mathcal{L}_{\mathrm{C22}}$ given by CHARMM22[33] and the MT/BLYP setup, we additionally computed a fourth set of 24×50 ps DFT/PMM-MD trajectories.

F. Optimizing van der Waals parameters for DFT/PMM

Given reference RDFs $g^{\mathrm{ref}}_{\mathrm{OH_w}}(r)$ and $g^{\mathrm{ref}}_{\mathrm{HO_w}}(r)$ for the solvation of NMA's AG by liquid water, which we obtained by first-principles DFT-MD as described in Section II D, the iterative DFT/PMM-MD search for optimal LJ parameters at the AG atoms of the DFT fragment is, in principle, straightforward.

1. Optimization scheme

For the optimization of LJ potentials at the AG atoms we assume that only the three heavy atoms $\gamma \in \{\mathrm{C}', \mathrm{O}, \mathrm{N}\}$ carry such potentials, whereas initially we adopt the CHARMM22 potentials[33] of the four AG atoms. This CHARMM22 parameter set is denoted $\mathcal{L}_{\mathrm{C22}}$. For the other atoms of the DFT fragment we always stick to the CHARMM22 potentials. Denoting the steps of the iterative search by $n = 1, 2, \ldots$, we thus probe in step n the LJ parameter set $\mathcal{L}_n = \{A^n_\gamma, B^n_\gamma \,|\, \gamma = \mathrm{C}', \mathrm{O}, \mathrm{N}\}$ for the LJ interaction of the GP6P oxygen atoms $\mathrm{O_w}$ and the AG atoms γ in the DFT fragment. Here, A^n_γ is the repulsion and B^n_γ the dispersion parameter of atom γ.

At each iteration step n one has to execute now, in principle, a sufficiently extended DFT/PMM-MD simulation, from which one can compute reliable estimates for the two

RDFs. Then the root $D(\mathcal{L}_n)$ of the summed mean square deviations between the reference and the DFT/PMM RDFs in the range $r \in [1.5, 2.5]$ Å, which covers the first peaks of these functions, can be used to measure the deviation of the hydrogen bonding structures in the water surrounding the AG computed by DFT/PMM from the reference. The parameters $\mathcal{L}_{n+1}$ for the next step in the minimization of $D(\mathcal{L})$ can then be chosen, e.g., by a simplex algorithm,[103] which eventually yields an optimized parameter set $\mathcal{L}_{\text{opt}}$ after sufficiently many iterations.

2. Mean-field approach employed for the optimization

The just described minimization of $D(\mathcal{L})$ poses a huge computational task, whose most expensive part is the DFT/PMM-MD simulations. In the spirit of the LJ parameter optimization presented by Martín *et al.*[104] we have, therefore, simplified these MD simulations by a mean-field method, which is similar to the so-called "averaged solvent electrostatic potential" approach[105–107] and which we denote as MFMD.

In our DFT/PMM-MFMD based optimization of the LJ parameters $\mathcal{L}$ we chose the MT/BLYP setup for the description of the NMA. As initial conditions of the 15 ps MFMD simulations executed at each iteration step n we chose the 24 phase space snapshots $\mathcal{S}_{\text{PMM}}$ introduced in Section II C.

In the MFMD simulations, the charge density ρ of the NMA surrounded by the GP6P water was self-consistently calculated for the initial snapshot, which requires the computation of the external potential $\Phi_0(\mathbf{r}_\gamma)$ generated by the PMM fragment at the points γ of the DFT grid.[10] This density and the positions of the NMA atoms were subsequently kept fixed in a 100 fs MD simulation, which solely involved the GP6P water molecules exposed to the potential generated by ρ_0. For the thus reached snapshot of the system, the external potential $\Phi_1(\mathbf{r}_\gamma)$ was recalculated and the average potential $\bar{\Phi}_1(\mathbf{r}_\gamma)$ was formed from this and the preceding snapshot $\Phi_0(\mathbf{r}_\gamma)$ of the potential.

(i) Then NMA's charge density ρ_1 was self-consistently computed for the average external potential $\bar{\Phi}_1(\mathbf{r}_\gamma)$ and the MD of the GP6P molecules exposed to the potential generated by ρ_1 was simulated for another 100 fs.
(ii) At this configuration the external potential $\Phi_2(\mathbf{r}_\gamma)$ and a running average potential $\bar{\Phi}_2(\mathbf{r}_\gamma)$ were calculated from $\Phi_2(\mathbf{r}_\gamma)$ and the preceding average potential $\bar{\Phi}_1(\mathbf{r}_\gamma)$.

Repeating the procedures (i) and (ii) a running average $\bar{\Phi}(\mathbf{r}_\gamma)$ of the external potential was sequentially accumulated for the 15 ps of MFMD simulation. RDFs $g_{\text{OH}_\text{w}}(r)$ and $g_{\text{HO}_\text{w}}(r)$ were extracted from the last 12 ps of all 24 MFMD simulations executed for the PMM liquid surrounding the NMA and the deviation $D(\mathcal{L}_n)$ of the current minimization step was calculated.

The just sketched DFT/PMM-MFMD approach saves about 99% of the computational effort associated with the import of the external potential on the DFT grid and with the computation of the Kohn-Sham orbitals, which are required at each time step of an ordinary DFT/PMM-MD simulation.

3. Calculation of vibrational spectra

From each 50 ps MD trajectory obtained in the DFT and DFT/PMM settings, respectively (cf. Sections II E 2 and II E 3) we calculated a corresponding IR spectrum of NMA by Fourier transforming the autocorrelation function of its dipole moment (FTTCF)[20] and by subsequently applying the so-called harmonic approximation quantum correction factor.[20,108,109] A Gaussian kernel of width 3 cm^{-1} was employed to smoothen the spectrum.[24] A final IR spectrum was obtained by averaging over the 12 (gas phase) or 24 (solution) trajectories associated to each setting. The bands in the IR spectra were assigned to vibrational modes by the generalized normal coordinate (GNC) analysis proposed by Mathias *et al.*,[110,111] which enables the decomposition of the vibrational spectrum into local modes directly from the MD trajectories.

III. RESULTS

As is indicated in Section II B and thoroughly explained in Section S1 of the supplementary material,[60] we have parameterized a new polarizable six-point model for water called GP6P specifically for the use as an aqueous solvent in DFT/PMM-MD simulations of solute molecules described by grid-based DFT. Before turning to the first applications, which are the parameterization of the LJ potentials steering the hydrogen bonding of NMA in GP6P water, and the subsequent calculation of NMA's IR spectrum in aqueous solution, we shortly sketch important microscopic and macroscopic properties of the GP6P water model.

A. The GP6P model

Table I lists the parameters of the GP6P model potential resulting from the DFT/PMM-based parameterization described in Section S1 of the supplementary material.[60] Figure 2 illustrates the thus determined electrostatic properties of GP6P.

1. Microscopic properties

Figure 2(a) explains the internal coordinates l_{OH}, φ_{HOH}, l_{OM}, l_{OL}, and φ_{LOL}, which define the rigid arrangement of the six points of force action. Here, the radii of the five colored

TABLE I. Parameters of the GP6P model.

l_{OH}/Å	0.968	φ_{HOH}/deg	105.3
μ_0/D	1.855	α_{PMM}/Å^3	1.47
l_{OL}/Å	0.595 528	φ_{LOL}/deg	177.360
l_{OM}/Å	0.558 801	q_{M}/e	−0.401 742
q_{H}/e	0.512 662	q_{L}/e	−0.311 791
σ_μ/Å	0.739	$\sigma_{\text{L}}/\sigma_{\text{H}}$	1.427 93
σ_{H}/Å	0.457 547	σ_{L}/Å = σ_{M}/Å	0.653 345
$A_1/(10^3$ kcal/mol)	11.241 6	A_2/Å^{-1}	2.875 96
B/(Å^6 kcal/mol)	966	$A/(10^3$ Å^6 kcal/mol)	518.941

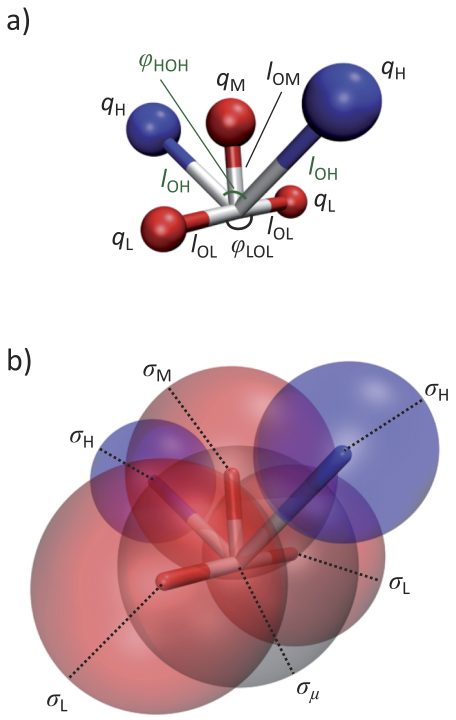


FIG. 2. (a) The geometry of the GP6P model is given by the parameters l_{OH}, φ_{HOH}, l_{OM}, l_{OL}, and φ_{LOL}, which define the HOH triangle of the water atoms and a perpendicular triangle LML of massless sites carrying charges $q_i < 0$ (red), $i \in \{M, L\}$. The H atoms harbor charges $q_H > 0$ (blue), the gray O atom is uncharged and carries an induced dipole $\boldsymbol{\mu}^i_{PMM}$. (b) All charges q_i and the induced dipole $\boldsymbol{\mu}^i_{PMM}$ form Gaussian distributions, whose widths σ_i, $i \in \{H, M, L, \mu\}$ are coded by the radii of the glassy spheres. Made with VMD.[95]

spheres code the sizes $|q_i|$ of the charges q_i, $i \in \{H,M,L\}$, while the uncharged oxygen is depicted in gray. Figure 2(b) illustrates the widths σ_i, $i \in \{H,M,L,\mu\}$, of the Gaussian charge (red, blue) and induced dipole (gray) distributions by the radii of the glassy spheres. The strength $\boldsymbol{\mu}^i_{PMM}$ of the induced dipole distribution centered at the oxygen atom follows by linear response from the polarizing electric field through the isotropic polarizability α_{PMM}.

As is indicated in Section II B, the polarizability α_{PMM}, the molecular geometry parameters (l_{OH}, φ_{HOH}), and the magnitude μ_0 of the vacuum dipole moment of GP6P have been adopted from experimental data. The associated values are listed in the first part of Table I. The central part of this table displays those parameters (including the ratio σ_L/σ_H of Gaussian charge widths), which were determined by DFT/PMM calculations. Finally, the bottom part of Table I specifies the parameters, which were determined by empirical PMM-MD optimizations. These are the Gaussian width σ_H and the Buckingham parameters (A_1, A_2, B), which describe the van der Waals interactions among the GP6P water molecules. Furthermore it provides the parameter A of a LJ potential [Eq. (S2) in the supplementary material[60]], which enables the description of van der Waals interactions between GP6P water and molecular models, whose atoms carry LJ potentials.

Note that Section S1 F 4 of the supplementary material[60] characterizes the dipole and quadrupole moments of a GP6P water model in the gas and liquid phases.

2. Macroscopic properties of GP6P at p_0 and T_0

Experimental values of the four macroscopic observables E_0, r_{OO}^{max}, p_0, and α_p served as targets for the empirical optimization of the four GP6P parameters σ_H, A_1, A_2, and B (cf. Section II B). The resulting parameter values lead to an excellent reproduction of the experimental target data. This claim is proven by Table S4 in the supplementary material,[60] which documents that the values computed from 3 ns PMM-MD simulations for the targeted observables numerically agree with their respective target values within the narrow limits of statistical accuracy. The oxygen-oxygen RDF $g_{OO}(r)$, which is shown by Figure S16 in the supplementary material[60] over the distance range $r \in [2,8]$ Å, furthermore reveals that the local ordering of water molecules predicted by GP6P resembles corresponding experimental data,[82] up to a slight over-structuring, quite well. Note that such a slight over-structuring is typical for RDFs derived from dynamics simulations, which neglect nuclear quantum effects.[112]

Table II lists GP6P predictions for several observables of liquid water, which were measured by PMM-MD simulations at the thermodynamic conditions (p_0,T_0) of the parameterization (our subsequent DFT/PMM-MD simulations of NMA in GP6P water will also be executed at these conditions). The predictions cover the density n, the isothermal compressibility κ_T, the heat capacity C_p, the self-diffusion constant D_0, the viscosity η, and the dielectric constant ϵ_0, which were calculated as explained in Section S1 D of the supplementary material.[60]

The density $n(p_0,T_0) = 0.9966 \pm 0.0001$ g/cm^3 resulting from a $N_m p_0 T_0$ simulation matches the experimental value $n^{exp}(p_0,T_0)$ within the narrow limits of statistical accuracy. This excellent match is not particularly surprising, because p_0 was one of the targets of the empirical optimization by $N_m V_m T_0$ simulations, in which the density had been set to $n^{exp}(p_0,T_0)$.

The isothermal compressibility $\kappa_T(p_0,T_0)$ determined for GP6P overestimates $\kappa_T^{exp}(p_0,T_0)$ by only 2.6%, whereas the heat-capacity $C_p(p_0,T_0)$ overestimates $C_p^{exp}(p_0,T_0)$ by 5.3%. The diffusion constant D_0 is by 25% larger than D_0^{exp} and the viscosity η happens to match η^{exp}. Finally the dielectric constant is by 10% smaller than the experimental reference. All these deviations of the GP6P predictions from the corresponding experimental values are comparable to those observed for standard MM water models[44] or for PMM

TABLE II. Observables for liquid water at 300 K: GP6P predictions and experimental data. Statistical errors of n, κ_T, and C_p were estimated by block averaging[113] from 3 ns trajectories, those of D_0, η, and ϵ from multiple and extended MD trajectories (see Sections S1 D and S1 F 2 of the supplementary material[60]).

Quantity	Unit	GP6P value	Expt.	Ref.
n	g/cm^3	0.9966 ± 0.0001	0.9965	80
κ_T	10^{-6}/atm	46.8 ± 0.1	45.6	80
C_p	cal/(mol K)	18.96 ± 0.06	18.0	114
D_0	nm^2/ns	3.01 ± 0.02	2.4	115
η	mPa s	0.81 ± 0.15	0.81	116
ϵ		69.9 ± 0.1	78	117

water models like TL4P[48] and BK3.[56] Thus, our aim of developing a PMM model potential for water, which describes the bulk properties of liquid water with a reasonable accuracy and, concurrently, solely exhibits Gaussian sources of the electrostatic potential, has certainly been reached.

3. The temperature-density profile of GP6P

The temperature dependence $n(p_0,T)$ of the density of liquid water has long been known at a very high accuracy. Kell[80] employed a ratio of fifth and first order polynomials [cf. Eq. (4) in Ref. 80] to fit his experimental data and thus determined the temperature $T_{\mathrm{md}}^{\mathrm{exp}} = 277.134$ K of maximum density and the maximal density $n^{\mathrm{exp}}(p_0,T_{\mathrm{md}}^{\mathrm{exp}}) = 0.999\,97$ g/cm^3 of liquid water at the standard pressure.

Figure 3 compares Kell's[80] experimental fit function $n^{\mathrm{exp}}(p_0,T)$ (gray) with the GP6P density profile $n(p_0,\langle T_k\rangle)$ (dots) extracted from a 20 ns replica exchange[118–120] MD simulation in the $N_{\mathrm{m}}p_0\tilde{T}_k$ ensembles (cf. Section S1 D[60]). Here, the temperatures $\langle T_k\rangle$ are averages over the trajectories of the various temperature rungs $\tilde{T}_k$. Moreover, the black line represents a fit to the simulation results $n(p_0,\langle T_k\rangle)$ using the same Ansatz for the fit function as Kell.[80]

Statistical errors of the calculated average densities $n(p_0,\langle T_k\rangle)$ were determined by block averaging.[113] The maximal error of 0.000 18 g/cm^3 was found for the most slowly converging ensemble at the lowest temperature rung $\tilde{T}_0 = 250$ K.

The fit to the GP6P data predicts the temperature T_{md} of maximum density at 277.469 ± 0.404 K and the maximum density $n(p_0,T_{\mathrm{md}})$ at 1.000 03 ± 0.000 04 g/cm^3. The statistical uncertainties of these numbers were estimated as the standard errors of the means, which were obtained by dividing the trajectories into four equal parts. Thus, the slight GP6P overestimate (0.335 K) of the experimental temperature[80] $T_{\mathrm{md}}^{\mathrm{exp}}$ of maximum density is within the statistical error, whereas the overestimate (0.000 06 g/cm^3) of the maximum density $n^{\mathrm{exp}}(p_0,T_{\mathrm{md}}^{\mathrm{exp}})$ is marginally larger.

The GP6P temperature-density profile in Figure 3 matches the experimental data[80] almost perfectly for temperatures $\langle T_k\rangle > T_{\mathrm{md}}$, whereas it slightly overestimates the experimental densities for lower temperatures $\langle T_k\rangle \ll T_{\mathrm{md}}$. The overall agreement is, however, excellent. This is pleasing particularly in the light of the fact that the empirical parameterization has targeted only observables at the standard conditions (p_0,T_0). Thus, the accurate prediction of $n(p_0,T)$ indicates that the GP6P model will most likely be quite well transferable to other conditions. Because the simulation software IPHIGENIE and sample GP6P simulation systems are available for download,[15] the scientific community is invited to carry out further studies on other issues of transferability.

When looking at Figure 3 the diligent reader may be reminded of Figure 3 in Ref. 121, which compares a density profile calculated by exactly the same replica exchange setup for a smaller box of TL6P models.[121] In fact, the deviations of the TL6P density profile[121] from the experimental data are only slightly larger than those obtained here with GP6P. However, because all TL6P simulations consistently employed an invalid barometer (cf. a corresponding remark in Section S1 B 1[60]) these results are seriously and consistently flawed, whereas the GP6P results are technically correct. On the other hand, the arrangement of the static partial charges in GP6P is similar to that of TL6P, because both models feature, beyond the negative charge at the M-site, two lone-pair charges at an angle φ_{LOL} close to 180° (cf. Fig. 2). It may well be that this arrangement[58] favors a density profile close to the experimental one, if the calculated thermal expansion coefficient $\alpha_p(p_0,T_0)$ is close to the experimental value.

This interpretation is supported by failed attempts (data not shown) to parameterize polarizable Gaussian four- and five-point models with the same strategy, which we successfully applied in the construction of GP6P. These attempts failed because no positive value for the dispersion parameter B could be found such that the isobaric thermal expansion coefficient α_p was close to the experimental target value. In this respect the Gaussian four- and five-point models resembled the polarizable TL4P and TL5P models, which feature point-like partial charges (and were parameterized and evaluated with a correct barometer) and both overestimate α_p by more than a factor of two.[48] Correspondingly their temperature-density profiles[121] show no maximum in the temperature range covered by Figure 3.

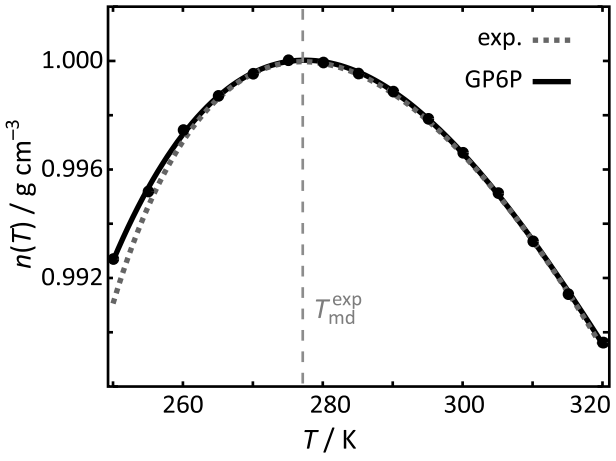


FIG. 3. Temperature-density profile of GP6P (dots) and a polynomial fit (black line). The experimental curve[80] is drawn in dashed gray.

B. LJ parameters for a DFT model of NMA

The optimization of the LJ parameters $\mathcal{L}$ (see Section II F), which steer the interactions of NMA's AG with the surrounding GP6P water, was conducted with the MT/BLYP setup for several hundred iteration steps. It reduced the observable $D(\mathcal{L})$ measuring the deviation of the DFT/PMM-MFMD RDFs $g_{\mathrm{OH_w}}(r)$ and $g_{\mathrm{HO_w}}(r)$ from their DFT-MD references by a factor of seven. Here, the initial deviation $D(\mathcal{L}_{\mathrm{C22}}) = 0.63$ Å, which resulted for the CHARMM22 parameter set $\mathcal{L}_{\mathrm{C22}}$, was reduced to $D(\mathcal{L}_{\mathrm{opt}}) = 0.09$ Å. Note that Figure S18 in the supplementary material[60] shows that the much cheaper DFT/PMM-MFMD calculations actually yield RDFs, which are almost identical to those obtained by costly DFT/PMM-MD simulations. The optimized set $\mathcal{L}_{\mathrm{opt}}$ resulting from DFT/PMM-MFMD is listed and compared with the

TABLE III. Optimized parameters $\mathcal{L}_{\text{opt}}$ for the LJ interaction of the AG atoms γ with the GP6P water model.

γ	$A_{\gamma,\text{opt}}$[a]	ΔA_γ (%)	$B_{\gamma,\text{opt}}$[b]	ΔB_γ (%)
C′	1747.34	−5	839.816	−7
O	451.865	+58	235.813	−36
N	924.673	−30	1306.43	+27
H	0.0	−100	0.0	−100

[a] 10^3 Å^{12} kcal/mol.
[b] Å^6 kcal/mol.

initial set $\mathcal{L}_{\text{C22}}$ in Table III. Here, the quantities ΔA_γ and ΔB_γ measure the optimization-induced parameter changes relative to $\mathcal{L}_{\text{C22}}$.

According to the listed data, the LJ potential at the central carbon atom C′ of NMA's AG remains almost invariant, whereas those at the O and N atoms experience sizable changes. Here, the O atom becomes much more repulsive and less attractive. Oppositely directed and somewhat weaker changes result for the N atom. The changes at the H atom follow from our deliberate removal of this very weak LJ potential.

Hence, one expects that the first peak of the RDF $g_{\text{OH}_\text{w}}(r)$, which is generated by water molecules hydrogen bonded to the C′═O group of NMA, substantially moves toward larger distances r upon the optimization. In contrast the peak of the RDF $g_{\text{HO}_\text{w}}(r)$, which describes the hydrogen bonding to the N—H group and, therefore, is mainly influenced by the parameters (A_N, B_N), should shift a little less toward smaller distances. The inspection of the black RDFs $g_{\text{OH}_\text{w}}(r)$ and $g_{\text{HO}_\text{w}}(r)$ shown in Figures 4(a) and 4(b), respectively, demonstrates that this is actually the case.

The dotted lines in Figure 4 mark the RDFs resulting from DFT/PMM-MD with the CHARMM22 parameter set $\mathcal{L}_{\text{C22}}$, whereas the solid black lines mark the RDFs obtained with $\mathcal{L}_{\text{opt}}$. In fact, the peak of the solid black curve $g_{\text{OH}_\text{w}}(r)$ in Figure 4(a) is found at a 7% larger distance than that of the associated dotted curve, whereas in Figure 4(b) the solid black peak of $g_{\text{HO}_\text{w}}(r)$ is found at a 3% smaller distance than the dotted peak. These shifts move the black curves much closer to the reference RDFs (solid gray), which were obtained by 85 ps first-principles DFT-MD. Note that these reference data carry non-negligible statistical uncertainties, which we estimated to be <1% for the locations and ≲6% for the heights of the peaks by partitioning the DFT-MD trajectory into four equal parts.

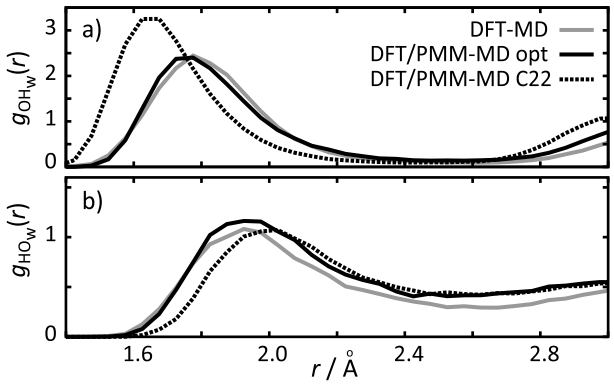


FIG. 4. RDFs (a) $g_{\text{OH}_\text{w}}(r)$ and (b) $g_{\text{HO}_\text{w}}(r)$ obtained from first-principles DFT-MD (solid gray), and from DFT/PMM-MD for the LJ parameter sets $\mathcal{L}_{\text{C22}}$ (dotted) and $\mathcal{L}_{\text{opt}}$ (solid black).

The RDFs calculated with $\mathcal{L}_{\text{opt}}$ (solid black) match the DFT references (solid gray) very well. In Figure 4(a) the DFT/PMM peak position deviates from that of the reference by only 1.4%, whereas in Figure 4(b) the corresponding peak positions show a perfect match. In the latter case DFT/PMM overestimates the height of the reference peak by 9%, whereas in the former case the heights are almost identical.

Now the interesting question arises as to whether the LJ parameter set $\mathcal{L}_{\text{opt}}$, which resulted from a DFT/PMM-MFMD optimization employing the MT/BLYP setup, can be generalized to other DFT setups. Figure S19 in the supplementary material[60] demonstrates that this is actually the case by comparing RDFs resulting from DFT/PMM-MD simulations with the MT/BLYP, MT/B3LYP, and MT/BP DFT setups, respectively. Thus, the LJ parameter set $\mathcal{L}_{\text{opt}}$ appears to be transferable to other DFT settings.

C. IR spectra of NMA from DFT-MD and DFT/PMM-MD simulations

IR frequencies calculated from DFT models usually require a scaling to achieve a match with experimental data.[41,42] Because we will derive the IR spectra of NMA in aqueous solution by the FTTCF technique[20] from DFT/PMM-MD simulations at the elevated temperature T_0 = 300 K, the NMA will dynamically sample also regions of the potential energy surface, which are shaped by anharmonicities. Our scaling factor should include these effects.

We will now first illustrate the effects of anharmonicities on the spectra of isolated NMA calculated by FTTCF from DFT-MD simulations at varying temperatures T. Subsequently we will compute scaling factors for three different DFT functionals by comparing experimental gas phase IR frequencies with results of DFT-MD simulations at T_0.

1. Anharmonicities in NMA's IR spectra obtained from DFT-MD

Because we wanted to gain insights, to what extent the IR spectra calculated from DFT-MD trajectories of isolated NMA at different temperatures are modified by anharmonicities, we have carried out 600 ps DFT-MD simulations at each of the target temperatures T/K = 20, 75, 150, 300, and 433 with the MT/BLYP approach (for details see Sections II E 1 and II E 2). Applying the FTTCF technique introduced in Section II F 3 to the trajectories $\boldsymbol{\mu}(t)$ of the molecular dipole moment yielded temperature dependent IR spectra, from which the spectral position $\nu_{\text{AI}}(T)$ of the dominant so-called amide-I (AI) band, which mainly belongs to the C′═O stretching vibration,[122] was deduced. Statistical errors $\sigma_\nu(T)$ of the frequencies $\nu_{\text{AI}}(T)$ were estimated as standard errors of the mean from the first and the second halves of the sets of DFT-MD trajectories.

Figure 5 shows the thus obtained frequencies $\nu_{\text{AI}}(\langle T\rangle)$ as a function of the actual average simulation temperatures $\langle T\rangle$ together with error bars $\sigma_\nu(\langle T\rangle)$ and a tentative linear

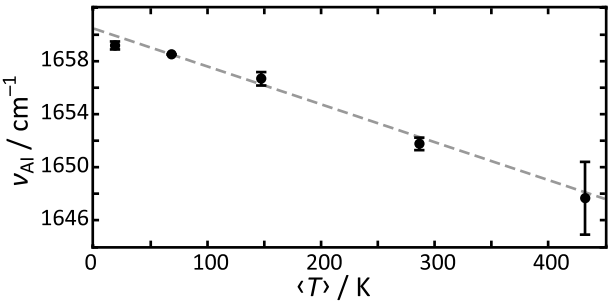


FIG. 5. Temperature dependence of the AI frequency ν_{AI} derived for isolated NMA by FTTCF (Ref. 20) from extended DFT-MD trajectories (setup: MT/BLYP). Error bars denote the standard errors σ_ν of the mean, the gray dashed line represents a linear regression.

regression line. One expects that less anharmonicities are sampled at low temperatures $\langle T\rangle$. Correspondingly the frequencies $\nu_{AI}(\langle T\rangle)$ are seen to increase almost linearly over a wide temperature range with decreasing $\langle T\rangle$ until they non-linearly level off for $\langle T\rangle \lesssim 75$ K. If one wants to choose a method-specific frequency scaling factor f_{DFT}, then one must be aware of this temperature dependence of ν_{AI} and should apply f_{DFT} only to simulations at the corresponding reference temperature, in our case T_0.

2. Scaling factors from DFT-MD at 300 K

For a more simple visual comparison with experimental spectra we choose a global frequency scaling, which moves the calculated gas phase frequency ν_{AI} exactly to the experimentally observed spectral position. The determination of the IR spectroscopic reference value encounters the difficulty that the AI band of NMA in the gas phase shows a distinct double peak,[29,123] which is caused by the coupling of the vibrations to the rotational degrees of freedom, while our computational setup excludes such a coupling. From the available gas phase IR data,[29] we estimate the value $\nu_{AI}^{exp} = 1722.5$ cm^{-1}. From our MT/BLYP-MD simulation result $\nu_{AI} = 1651.8$ cm^{-1}, which belongs to the average simulation temperature $\langle T\rangle = 287$ K, we get the scaling factor $f_{MT/BLYP} = 1.0428$.

A priori it is uncertain, whether gas phase scaling factors f_{DFT} (like $f_{MT/BLYP}$) are applicable also to liquid phase IR spectra derived from DFT/PMM-MD simulations of NMA in GP6P water at $T_0 = 300$ K. While the temperature dependence of the calculated anharmonic frequencies should be well-accounted for by f_{DFT}, the transition to a hybrid setting implies a change of the computational method. Therefore it may well be that the liquid phase IR spectra require a slightly modified scaling factor f_{DFT}^{sol} for a perfect match of ν_{AI}^{sol} with the associated $\nu_{AI}^{sol,exp}$. This factor should then be applicable to IR spectra calculated by DFT/PMM-MD for other peptides in GP6P water.

For the MT/BP setup, an analogous DFT-MD simulation of NMA in the vacuum at the average temperature $\langle T\rangle = 272$ K yielded the smaller scaling factor $f_{MT/BP} = 1.033\,45$. This factor is very close the value of 1.0354, which had been determined in an earlier MT/BP study on NMA[24] by comparing the results of a normal mode analysis with the frequency $\nu_{AI}^{RR} = 1728$ cm^{-1} measured by resonance Raman spectroscopy.[124]

Interestingly, the much more costly (cf. Section III D) MT/B3LYP-MD simulations, which were carried out at an average temperature $\langle T\rangle$ of 327 K, yielded a scaling factor $f_{MT/B3LYP} = 0.9988$ very close to one. Thus, the inclusion of Hartree-Fock exchange[125] substantially improves and stiffens the BLYP force field of NMA in the region of the energy surface sampled by our DFT-MD. Therefore we will now discuss the IR spectra of NMA in the gas phase and in aqueous solution on the basis of the MT/B3LYP simulation results and will consider the effects of the chosen DFT setup on the quality of the predicted IR spectra further below.

3. The IR spectrum of NMA in the gas phase

Applying the just determined scaling factor $f_{MT/B3LYP}$ to the gas phase IR spectrum of NMA calculated by FTTCF from the 600 ps MT/B3LYP-MD trajectories (cf. Section II E 2) we obtained the spectrum depicted in Figure 6(a) as a solid black line. The thus calculated IR bands were assigned to the various amide modes (cf. Figure 7 in Ref. 24) by a GNC analysis.[110,111] The gray spectrum in the background has been measured by IR spectroscopy.[29]

The spectral locations of the calculated amide peaks match the positions of the corresponding observed bands very well. For the AI band this frequency match is, of course, a result of the scaling. In contrast, the spectral positions of the AII, AIII, and AIV bands calculated at 1492 cm^{-1}, 1248 cm^{-1}, and 1071 cm^{-1}, respectively, represent predictions. Here, the excellent match of the calculated AII and AIII peak frequencies with their experimental counterparts (AII: 1499 cm^{-1}, AIII: 1255 cm^{-1}), which can be quantified by

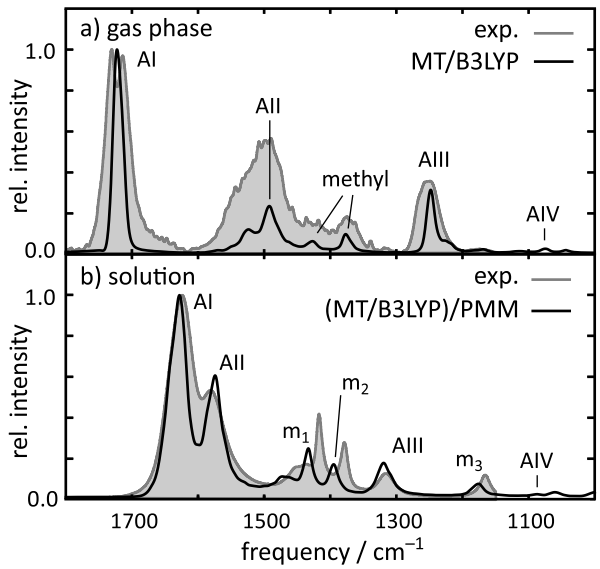


FIG. 6. The IR spectrum of NMA in (a) the gas phase and (b) aqueous solution as found by IR spectroscopy[29] (gray) and by FTTCF (black) from DFT-MD and DFT/PMM-MD trajectories employing the MT/B3LYP setup; frequencies were scaled with $f_{MT/B3LYP}$; the heights of the AI peaks were normalized to one; band assignments were obtained by a GNC analysis.[110,111]

deviations of at most 7 cm^{-1}, demonstrates the quality of the marginally scaled MT/B3LYP-MD description. Because there are no IR data in the frequency range below 1150 cm^{-1} available to us, we cannot judge the accuracy of the predicted AIV frequency.

Figure S20(a) in the supplementary material[60] furthermore demonstrates that, for the amide bands, the quality of the results depends (after proper scaling) hardly on the chosen DFT setup. This figure shows that MT/BLYP-MD and MT/BP-MD simulations lead to likewise excellent description of these bands as the MT/B3LYP-MD simulations. Interestingly, the marginally scaled MT/B3LYP-MD describes the methyl bands at about 1428 cm^{-1} and 1375 cm^{-1} distinctly better than scaled MT/BLYP-MD and at a quality comparable to that of scaled MT/BP-MD.

The above analysis in terms of band frequencies solely confirms the visual comparison of the calculated with the underlying experimental IR spectrum, which clearly indicates an almost perfect frequency match. The comparison of peak heights is not likewise perfect. This is partially due to the fact that the applied normalization of the AI peak heights represents a poor intensity measure. Intensities are actually integrals over bands, which are poorly measured by the peak heights. Furthermore, the artificial suppression of the molecular rotations in the DFT-MD simulations, which is enforced by the lacking conservation of the angular momentum with a grid-based DFT implementation like in CPMD,[14] entails an artificial narrowing of the various bands. Particularly in the case of the AI band, this suppression transforms the observed vibrational-rotational double peak into a single one. Therefore, the relative intensity of the AI band is strongly overestimated by the applied peak height normalization. Despite the resulting difficulty of intensity comparisons, the computational results enable together with the GNC analysis a clear-cut assignment of observed bands to molecular modes.

4. The IR spectrum of NMA in aqueous solution

Now it will be interesting to see as to whether IR spectra, which are calculated from our 1.2 ns DFT/PMM-MD trajectories of NMA in GP6P water at 300 K (cf. Section II E 3) using the optimized LJ parameter set $\mathcal{L}_{opt}$ for the AG atoms of NMA, provide a similarly excellent description of spectroscopic IR data on NMA in liquid water at room temperature. Figure 6(b) compares the slightly scaled FTTCF results obtained with the MT/B3LYP setup (black) with a corresponding experimental IR spectrum[29] (gray).

Already the first glance at this figure shows that the FTTCF approach predicts the three amide bands also in the solvent case at spectral locations, which are very close to those determined by IR spectroscopy.[29] For the shown six prominent IR peaks the root mean square deviation (RMSD) amounts to 9 cm^{-1}. The dominant contributions to this RMSD are delivered by the methyl bands (denoted as m_1, m_2, and m_3), which are on average blue-shifted by 12 cm^{-1} with respect to the spectroscopic band positions. The three amide bands AI-AIII, for which experimental reference data are available, contribute only a RMSD of 5 cm^{-1}.

Overall the calculated spectrum reproduces observed spectral features very well although it cannot account,[19] because of the use of a rigid PMM water model, for the likely coupling[126,127] of the AI vibration with the bending mode of H_2O at 1644 cm^{-1},[128] which may induce a broadening and an additional shift of the AI band. In combination with the GNC analysis[110,111] the calculations therefore enable a clear-cut assignment of the observed bands to local modes. The AIV band is predicted, e.g., at 1085 cm^{-1}.

Combining the results in Figures 6(a) and 6(b) on the amide bands of NMA in the gas phase and in solution we conclude that the experimentally determined[29] solvatochromic shifts of the AI, AII, and AIII bands, which are −98 cm^{-1}, +83 cm^{-1}, and +62 cm^{-1}, respectively, are reproduced by our DFT/PMM-MD approach in the marginally scaled MT/B3LYP setup with a RMSD of only 6 cm^{-1}. For the gas and liquid phases the bottom of Figure S21 in the supplementary material[60] provides further MT/B3LYP predictions on the frequencies of the AIV, AV, and AVI bands, which appear below 1100 cm^{-1}. Here its is shown that these amide bands are blue-shifted by 14 cm^{-1}, 41 cm^{-1}, and 20 cm^{-1}, respectively, upon transfer of NMA from the vacuum into GP6P water.

As a first physical result we note that the above DFT-MD and DFT/PMM-MD descriptions of the IR spectra of NMA in the gas phase and in aqueous solution predict the solvatochromic IR band shifts at a high accuracy. Concerning the methodology we furthermore note, that our initial assumption on the transferability of frequency scaling factors from a pure DFT to a hybrid DFT/PMM setting seems to be approximately valid, at least for the MT/B3LYP setup.

5. Effect of DFT functional on the IR spectra of NMA in GP6P water

Due to the much larger computational cost (cf. Section III D) of the B3LYP functional one might prefer the simpler gradient-corrected BLYP or BP functionals for the computation of IR spectra by MD-based FTTCF techniques. Further above in Section III C 3, we have noted that DFT-MD yields for NMA in the gas phase scaled MT/BLYP and MT/BP amide band frequencies, which are of a comparable quality to those obtained by the marginally scaled MT/B3LYP-MD. Here the question arises, whether this favorable result remains valid in DFT/PMM-MD calculations of liquid phase IR spectra.

Figure 7 now demonstrates that this is actually the case for NMA in aqueous solution. Here, the visual comparison of the amide frequency terms in the last two columns illustrates the small 5 cm^{-1} RMSD between the experimental frequencies[29] and the DFT/PMM-MD predictions obtained through the marginally scaled MT/B3LYP setup [cf. the discussion of Figure 6(b) above]. Following the corresponding predictions toward the first two columns of the figure, which belong to the scaled MT/BP and MT/BLYP setups, one recognizes that the qualities of the amide band predictions deteriorate only a little as signified by RMSDs of 8 cm^{-1} and 9 cm^{-1}, respectively. For the three methyl bands the situation is different. While their frequencies are blue-shifted with respect to the experimental data on the average by only 14 cm^{-1} for the scaled MT/BP setup, which is very close to the MT/B3LYP blueshift of

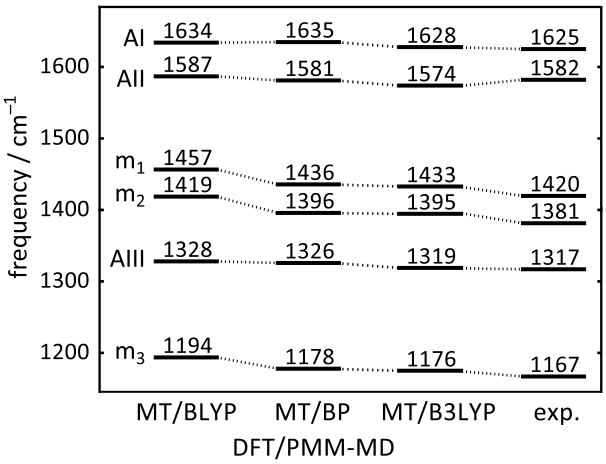


FIG. 7. NMA in aqueous solution: frequencies of amide (AI-AIII) and methyl (m_1-m_3) IR bands predicted by DFT/PMM-MD with the scaled MT/BLYP, MT/BP, and MT/B3LYP setups are compared with experimental data.[29]

12 cm^{-1}, a much larger average blueshift of 34 cm^{-1} results for the scaled MT/BLYP setup.

The residual 25 cm^{-1} and 11 cm^{-1} blueshifts remaining for the six IR bands in the scaled MT/BLYP and MT/BP setups, respectively, indicate that the transfer of scaling factors from vacuum DFT-MD to hybrid DFT/PMM-MD simulations of solute molecules may be suboptimal for these DFT setups. Thus, our tentative conclusion at the bottom of Section III C 4 on the transferability of the scaling factor seems to be restricted to the MT/B3LYP case. Note that Sections S3 A and S3 B of the supplementary material[60] present the IR spectra of NMA in the gas and aqueous phases computed for the MT/BLYP and MT/BP settings in the ranges [1800, 1000] cm^{-1} and [1380, 580] cm^{-1}, respectively.

The MT/BLYP and MT/BP band frequencies shown in Figure 7 for NMA in GP6P water suggest that one should choose for IR spectra of organic solute molecules derived from such DFT/PMM-MD simulations the scaling factors $f^{sol}_{MT/BLYP} = 1.0371$ and $f^{sol}_{MT/BP} = 1.0273$, which shift the calculated AI frequencies ν^{sol}_{AI} to the experimental value[29] $\nu^{sol,exp}_{AI} = 1625$ cm^{-1}. With these scaling factors the RMSD of the five remaining IR frequencies shown in Figure 7 reduces from 27 cm^{-1} to 21 cm^{-1} (MT/BLYP) and from 12 cm^{-1} to 7 cm^{-1} (MT/BP), respectively. The latter RMSD is even smaller than the 9 cm^{-1} RMSD resulting for the six main IR peaks in the marginally scaled MT/B3LYP case.

Thus, a properly scaled MT/BP setup can predict the IR spectra of NMA at a quality comparable to that of the physically more accurate and much more costly MT/B3LYP setup. This finding reminds of earlier observations derived from normal mode analyses[41,42] of isolated organic molecules, according to which unscaled harmonic BP frequencies match for many organic molecules the frequencies of anharmonic fundamentals surprisingly well. The associated accidental cancellation of errors[41,42] thus appears to transfer to the DFT/PMM-MD setting. Note in this context that MD simulations at about 300 K still leave a molecule in regions of the potential energy surface, which are quite close to the quadratic minimum.

6. Effects of NMA's LJ parameters on calculated IR spectra

In Section III B, we have shown that the solvation of NMA's C′=O group by GP6P water is substantially overestimated, if the LJ parameter set $\mathcal{L}_{C22}$, which is provided by the CHARMM22 force field for NMA, is used for the AG atoms in DFT/PMM simulations. In this respect the optimized set $\mathcal{L}_{opt}$ was seen to perform much better (cf. Figure 4).

Because the AI and AII bands of NMA are very sensitive to the strength of hydrogen bonding particularly at the C′=O group of the AG,[24,29,32] one expects that the AI band is shifted toward higher and the AII band toward lower frequencies if $\mathcal{L}_{C22}$ is replaced by $\mathcal{L}_{opt}$ in DFT/PMM-MD simulations.

The top part of Figure 8 confirms the expectation voiced above: Due to the exchange of the initial LJ parameter set $\mathcal{L}_{C22}$ by $\mathcal{L}_{opt}$ the AI band gets blue-shifted by 8 cm^{-1} and the AI band red-shifted by 7 cm^{-1} thus widening the spectral gap between the two bands from 32 cm^{-1} to 47 cm^{-1}. Here the experimental value[29] is 43 cm^{-1}.

Turning now to the AIII frequency shown in the bottom part of Figure 8 one recognizes a shifting pattern, which is quite similar to that of the AII frequency. Also here the transition from the $\mathcal{L}_{C22}$ to the $\mathcal{L}_{opt}$ parameters induces a redshift and, thus, diminishes the frequency overestimate resulting for the scaled DFT/PMM-MD description with the MT/BLYP setup (we have discussed the remaining frequency overestimates above already in connection with Figure 7). Note that the methyl bands are unaffected by the change of $\mathcal{L}$.

As a further methodological result it has thus become clear that a more accurate description of NMA's solvation by optimized LJ parameters at the two molecular groups, which are targeted by hydrogen bonding, leads to an improved prediction on the IR spectrum of NMA in aqueous solution. Therefore we suggest to adopt for those QM atoms, which are involved in hydrogen bonding interactions, an RDF-based LJ parameter optimization as a standard technique for the setup of QM/MM descriptions.

An alternative and less costly approach to the optimization of QM/MM van der Waals interaction potentials applies microsolvation models,[34–38] i.e., small isolated solute-solvent

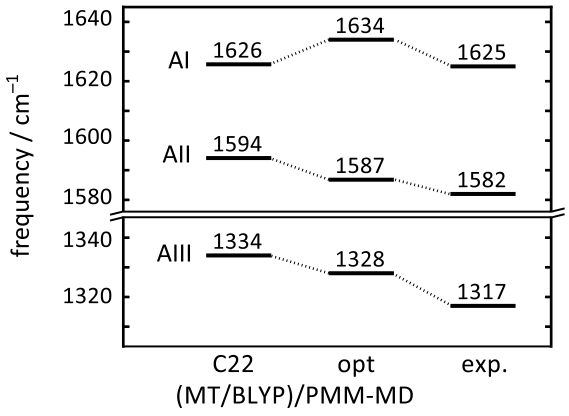


FIG. 8. Frequencies of the AI, AII, and AIII bands from DFT/PMM-MD in the scaled MT/BLYP setup either using the CHARMM22 ($\mathcal{L}_{C22}$) or the optimized ($\mathcal{L}_{opt}$) LJ parameters for the AG of NMA. As an experimental reference also data from IR spectroscopy[29] are given.

clusters, for which accurate QM reference calculations are feasible. However, the transfer of LJ parameters to the bulk failed in certain cases.[129]

7. Basic vs. advanced MD simulation technology and modeling

About eight years ago, Schultheis *et al.*[24] have already calculated the IR spectrum of NMA in water from DFT/MM-MD simulations. For the DFT fragment NMA they had chosen the MT/BP setup and for the MM description of the surrounding water Jorgensen's so-called TIP4P[43] potential. Beyond exchanging the non-polarizable and less polar TIP4P water model (dipole moment 2.18 D) by the polarizable and more polar GP6P description (average dipole moment in bulk water 2.54 D, cf. Table S5 in the supplementary material[60]), we have replaced the DFT/MM interface originally developed by Eichinger *et al.*[8] by improved algorithms, which substantially enhance the performance (i.e., accuracy times efficiency) of the calculations.[10,11,17] In this work, we furthermore have optimized the LJ parameters of NMA for an improved description of the hydrogen bonding structures at the C′═O and N—H groups of the AG.

Figure 9 compares the IR spectra calculated in the MT/BP setup from hybrid MD simulations with the advanced (a) and basic (b) modeling technologies, respectively. Whereas the description of the methyl bands is of a comparable quality in both cases, the amide bands resulting from the basic modeling are blue-shifted with respect to the advanced modeling by 18 cm^{-1} (AI), 15 cm^{-1} (AII), and 5 cm^{-1} (AIII). Whereas the blueshift of the AI band could be attributed to the use of the less polar TIP4P water model, a reduced polarity of the solvent should correspondingly lead to a redshift of the AII and AIII bands (these are the well-known effects exerted

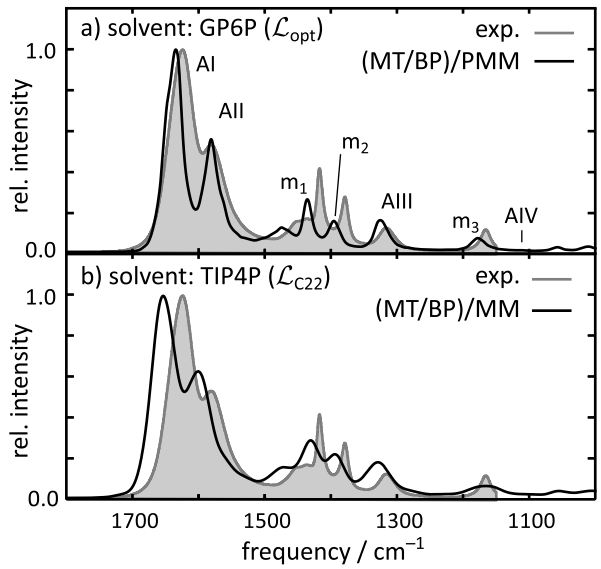


FIG. 9. The IR spectrum of NMA computed by FTTCF (black) from (a) DFT/PMM-MD with the GP6P water model and the LJ parameter set $\mathcal{L}_{opt}$, or from (b) DFT/MM-MD with the TIP4P[43] water model and $\mathcal{L}_{C22}$ (both employing the MT/BP setup). The AI intensities have been normalized to one and the frequencies were scaled by $f_{MT/BP}$. As a guide to the eye the experimental spectrum[29] is drawn in gray.

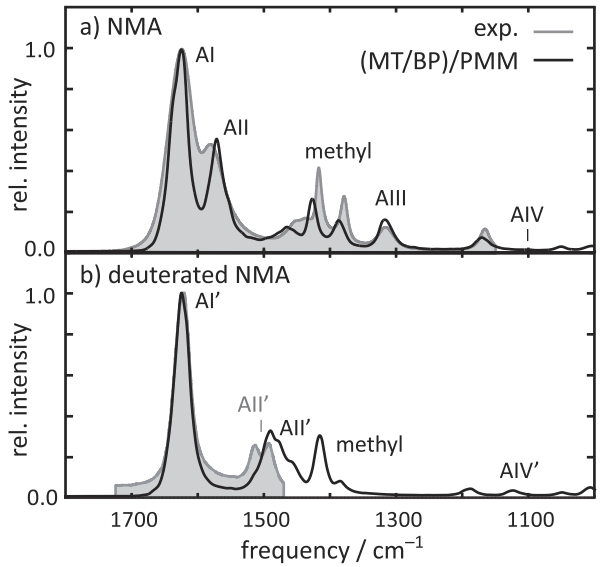


FIG. 10. IR spectra (black) of (a) NMA and (b) deuterated NMA computed by FTTCF from MT/BP DFT/PMM-MD and scaled by $f^{sol}_{MT/BP}$ are compared with available spectroscopic reference data[29] (gray). Band assignments were obtained by a GNC analysis.[110,111]

by solvents of reduced polarity[24,29,32]). Thus, the observed blueshift of the AII and AIII bands cannot be attributed to the smaller polarity of the TIP4P solvent. As a result, the improved description of the experimental data by the advanced DFT/PMM-MD simulations signifies the meanwhile achieved progress of technology,[10,11] methodology, and modeling.

8. Isotope effects: Deuteration of amide hydrogen

As a first application of the thus established DFT/PMM-MD approach to the IR spectra of AG-containing compounds in solution we consider the effect of replacing the amide hydrogen of NMA by a deuterium, because this is the only isotopomer of NMA, for which a solution spectrum[29] is known to us.

Figure 10 compares our DFT/PMM-MD predictions (black) for the IR spectra of NMA (a) and of its deuterated counterpart (b) with spectroscopic data.[29] Here, the MT/BP setting was employed and the associated *solution* frequency scaling factor $f^{sol}_{MT/BP}$ was applied (cf. Section III C 5). Unfortunately the experimental spectrum covers only a limited spectral range, because the deuterated aqueous solvent exhibits strong IR absorptions in the lower frequency spectral region.[128] Nevertheless the shown spectroscopic data[29] suffice to estimate the isotopic frequency shifts of NMA's AI and AII bands in the transition to deuterated NMA, where they are labeled as AI′ and AII′, respectively. The two bands are red-shifted by 2 cm^{-1} (AI) and 79 cm^{-1} (AII). In excellent agreement with these spectroscopic data our DFT/PMM-MD calculations predict redshifts of 1 cm^{-1} and 81 cm^{-1}, respectively, underlining the quality of the description.

D. Computational issues

The simulations were carried out on the SuperMUC Phase 2 Petascale System at the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities

in Munich employing up to 840 cores per simulation. The DFT fragment was handled by the MPI/OpenMP-parallelized DFT program CPMD[14] and the PMM fragment by the MPI-parallelized dynamics driver IPHIGENIE.[15] Here, CPMD works as a library loaded by IPHIGENIE.[11]

For MT/BP and MT/BLYP, the fastest parallel setup yielded about 70 ps of DFT/PMM-MD trajectory of NMA in GP6P solution per day, which is only by a factor of two slower than the corresponding DFT-MD simulations of isolated NMA. For DFT-MD simulations of isolated NMA, the MT/B3LYP setup was by about a factor of eight slower than the MT/BLYP setup, whereas in the DFT/PMM case this factor reduced to five. Nevertheless, applications employing the B3LYP functional with the current DFT/PMM implementation[11] are likely too expensive in many cases. Alleviation could provide a DFT/PMM interface to a more efficient B3LYP implementation as it is, e.g., available in CP2K.[63] The programming of such an extension should be straightforward.

With IPHIGENIE single-core test simulations of a box of liquid water containing 1500 molecules revealed that the new GP6P potential is only by 7.5% slower than TL6P,[49] whose model complexity differs from that of GP6P solely by the use of point instead of Gaussian partial charges.[59] Thus, the additional cost of handling Gaussian partial charge distributions is very small already in pure PMM-MD simulations and becomes essentially negligible in a DFT/PMM-MD setting.

IV. SUMMARY

A physically adequate and quantitatively reliable treatment of the quantum-classical interface in MD simulations of QM/MM hybrid systems still poses technical challenges, even if only non-covalently bound QM and MM fragments are in the focus, as it is the case in our present contribution. One of these challenges is the question, as to how one should describe the van der Waals interactions between the atoms of charged or polar fragments.

Driven by our interest in the IR spectra of peptides in their native aqueous environment, we have selected the standard molecular model NMA of a peptide bond and, here, the effects of solvation on its IR spectrum as the target of our study. Because the IR spectra of NMA, in particular, and of peptides, in general, sensitively depend[24,29,32] on the details of the electrostatic interactions of these strongly polar and polarizable molecules with their likewise polar and polarizable (native) environments, these environments should be modeled by PMM potential functions, which accurately capture their electrostatic properties.

Applying a DFT/PMM hybrid method[10,11] we have correspondingly developed with GP6P a polarizable model potential for water, which should be specifically well-suited for DFT/PMM-MD simulations of peptides in aqueous solution, if the peptidic DFT fragment is described by a grid-based DFT approach such as provided by CPMD[14] or CP2K.[63] Here, the Gaussian charge and dipole distributions of GP6P guarantee that spurious distortions of the DFT fragment's electron density, which may arise from point sources of the electrostatic potential, are avoided. Point charge MM models, in contrast, require additional efforts of selecting suitable charge smoothing scales.[9] The thus obtained GP6P model turned out to have favorable bulk liquid properties not only at the thermodynamic conditions (p_0, T_0) of its parameterization but also seems to be very well transferable to other liquid phase conditions.

Subsequently we have addressed the issue of the inter-fragment van der Waals interactions specifically for the two hydrogen bonding interaction sites of DFT-described NMA in GP6P water. We have derived for these sites an optimized set $\mathcal{L}_{\text{opt}}$ of LJ parameters by consideration of RDFs measuring the hydrogen bonding structure in the surrounding GP6P water. Reference RDFs were calculated by a first-principles DFT-MD simulation of NMA in a small box of water. The computational effort of the DFT/PMM-MD parameter optimization was reduced by applying a mean field approach.

We have evaluated the successes of these modeling efforts, i.e., the computational constructions of GP6P and of $\mathcal{L}_{\text{opt}}$, by computing the IR spectrum of NMA in aqueous solution through FTTCF from extended DFT/PMM-MD simulations using several DFT setups. The results turned out to represent excellent descriptions of the corresponding spectroscopic evidence[29] particularly for the marginally scaled MT/B3LYP and the scaled MT/BP setups. Here, the required scaling factors had been derived by comparing experimental gas phase IR data with results of DFT-MD simulations. The quality of the thus achieved DFT/PMM-MD description was underlined by the fact that the observed effects[29] of aqueous solvation and of deuteration on the IR bands of NMA were closely reproduced by the simulations.

The contents of this contribution are admittedly quite technical. An exception is, of course, the substantially improved description of the IR spectra of protonated and deuterated NMA in water. However, the achieved technical progress now provides solid grounds for DFT/PMM-MD studies of larger peptides in aqueous solution. Of particular interest for us will be the conformational landscapes and the IR spectra of dipeptides in aqueous solution. An enhanced sampling method, which makes DFT/PMM-MD simulations aiming at conformational landscapes computationally feasible, has been recently developed.[130]

The improved methodology of DFT/PMM simulations presented in this and previous works[10,11] is, up to now, restricted to chemically separated DFT and PMM fragments. An extension toward covalently connected fragments should pose no serious conceptual difficulties, if one follows the lines of reasoning underlying the so-called "scaled position link atom" method.[8]

ACKNOWLEDGMENTS

Computer time was provided by the Leibniz-Rechenzentrum (No. PR89XE); financial support by the Deutsche Forschungsgemeinschaft (Grant No. SFB749/C4) and the "Bavarian Competence Network for Technical and Scientific High Performance Computing" (KONWIHR-III) is acknowledged. The authors thank Konstantin Lorenzen for technical support.

[1]A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[2]H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[3]E. Brunk and U. Rothlisberger, Chem. Rev. **115**, 6217 (2015).
[4]M. Schwörer and G. Mathias, "Modeling biomolecular solvation effects by hybrid QM/MM methods," in *Computational Trends in Solvation and Transport in Liquids*, edited by G. Sutmann, J. Grotendorst, G. Gompper, and D. Marx, Schriften des Forschungszentrums Jülich, IAS Series Vol. 28 (Forschungszentrums Jülich, 2015), pp. 315–332.
[5]R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[6]M. P. Gaigeot, R. Vuilleumier, M. Sprik, and D. Borgis, J. Chem. Theory Comput. **1**, 772 (2005).
[7]J. VandeVondele, P. Tröster, P. Tavan, and G. Mathias, J. Phys. Chem. A **116**, 2466 (2012).
[8]M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello, J. Chem. Phys. **110**, 10452 (1999).
[9]A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
[10]M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan, and G. Mathias, J. Chem. Phys. **138**, 244103 (2013).
[11]M. Schwörer, K. Lorenzen, G. Mathias, and P. Tavan, J. Chem. Phys. **142**, 104108 (2015).
[12]P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[13]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[14]J. Hutter, A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman, and M. Parrinello, CPMD: Car–Parinello Molecular Dynamics, version 3.17.1, ©IBM Corp. 1990–2008 and MPI für Festkörperforschung Stuttgart 1997–2001, www.cpmd.org.
[15]IPHIGENIE is available for download free of charge under the GPL licence at http://sourceforge.net/projects/iphigenie.
[16]K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, and P. Tavan, J. Chem. Theory Comput. **8**, 3628 (2012).
[17]K. Lorenzen, C. Wichmann, and P. Tavan, J. Chem. Theory Comput. **10**, 3244 (2014).
[18]K. Lorenzen, G. Mathias, and P. Tavan, J. Chem. Phys. **143**, 184114 (2015).
[19]M. Schmitz and P. Tavan, J. Chem. Phys. **121**, 12233 (2004).
[20]M. Schmitz and P. Tavan, J. Chem. Phys. **121**, 12247 (2004).
[21]M. Schmitz and P. Tavan, "On the art of computing the IR spectra of molecules in condensed phase," in *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, edited by E. Starikov, S. Tanaka, and J. Lewis (Elsevier, Amsterdam, 2006), Chap. 8, pp. 157–177.
[22]J. Gao and M. Freindorf, J. Phys. Chem. A **101**, 3182 (1997).
[23]S. Yang and M. Cho, J. Chem. Phys. **123**, 134503 (2005).
[24]V. Schultheis, R. Reichold, B. Schropp, and P. Tavan, J. Phys. Chem. B **112**, 12217 (2008).
[25]J. Jeon and M. Cho, New J. Phys. **12**, 065001 (2010).
[26]F. Ingrosso, G. Monard, M. Hamdi Farag, A. Bastida, and M. F. Ruiz-López, J. Chem. Theory Comput. **7**, 1840 (2011).
[27]M. H. Farag, A. Bastida, M. F. Ruiz-Lopez, G. Monard, and F. Ingrosso, J. Phys. Chem. B **118**, 6186 (2014).
[28]N. G. Mirkin and S. Krimm, J. Mol. Struct. **377**, 219 (1996).
[29]J. Kubelka and T. A. Keiderling, J. Phys. Chem. A **105**, 10922 (2001).
[30]B. Mennucci and J. M. Martínez, J. Phys. Chem. B **109**, 9818 (2005).
[31]M. H. Farag, M. F. Ruiz-López, A. Bastida, G. Monard, and F. Ingrosso, J. Phys. Chem. B **119**, 9056 (2015).
[32]N. A. Besley, J. Phys. Chem. A **108**, 10794 (2004).
[33]A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[34]P. Bash, L. Ho, A. D. MacKerell, Jr., D. Levine, and P. Hallstrom, Proc. Natl. Acad. Sci. U. S. A. **93**, 3698 (1996).
[35]M. Freindorf and J. Gao, J. Comput. Chem. **17**, 386 (1996).
[36]F. Luque, N. Reuter, A. Cartier, and M. Ruiz-López, J. Phys. Chem. A **104**, 10923 (2000).
[37]D. Riccardi, G. Li, and Q. Cui, J. Phys. Chem. B **108**, 6467 (2004).
[38]U. Pentikäinen, K. E. Shaw, K. Senthilkumar, C. J. Woods, and A. J. Mulholland, J. Chem. Theory Comput. **5**, 396 (2009).
[39]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
[40]D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, 2009).
[41]M. Nonella and P. Tavan, Chem. Phys. **199**, 19 (1995).
[42]J. Neugebauer and B. A. Hess, J. Chem. Phys. **118**, 7215 (2003).
[43]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[44]C. Vega and J. L. Abascal, Phys. Chem. Chem. Phys. **13**, 19663 (2011).
[45]M. Klähn, G. Mathias, C. Kotting, J. Schlitter, M. Nonella, K. Gerwert, and P. Tavan, J. Phys. Chem. A **108**, 6186 (2004).
[46]G. Babitzki, G. Mathias, and P. Tavan, J. Phys. Chem. B **113**, 10496 (2009).
[47]B. Rieff, S. Bauer, G. Mathias, and P. Tavan, J. Phys. Chem. B **115**, 11239 (2011).
[48]P. Tröster, K. Lorenzen, M. Schwörer, and P. Tavan, J. Phys. Chem. B **117**, 9486 (2013).
[49]P. Tröster, K. Lorenzen, and P. Tavan, J. Phys. Chem. B **118**, 1589 (2014).
[50]P. Hamm, J. Chem. Phys. **141**, 184201 (2014).
[51]Unphysical distortions of the electron density due to the use of point charges in the MM fragment occur also in QM/MM calculations, which employ Gaussian basis functions in the DFT fragment; for a discussion of possible remedies see Refs. 52 and 53.
[52]S. Chalmet and M. F. Ruiz-Lopez, Chem. Phys. Lett. **329**, 154 (2000).
[53]Y. Jin, E. R. Johnson, X. Hu, W. Yang, and H. Hu, J. Comput. Chem. **34**, 2380 (2013).
[54]P. K. Biswas and V. Gogonea, J. Chem. Phys. **123**, 164114 (2005).
[55]T. Laino, F. Mohamed, A. Laio, and M. Parrinello, J. Chem. Theory Comput. **1**, 1176 (2005).
[56]P. T. Kiss and A. Baranyai, J. Chem. Phys. **138**, 204507 (2013).
[57]P. Paricaud, M. Předota, A. Chialvo, and P. Cummings, J. Chem. Phys. **122**, 244511 (2005).
[58]M.-L. Tan, J. R. Cendagorta, and T. Ichiye, J. Chem. Phys. **141**, 244504 (2014).
[59]For instance, with SAMM the computational overhead of the polarizable six-point model TL6P[49] over the non-polarizable three-point model TIP3P[43] is only a factor of 4.6 although, in the case of the polarizable model, the induced dipoles have to be self-consistently evaluated at each integration step of a MD simulation.[17]
[60]See supplementary material at http://dx.doi.org/10.1063/1.4943972 for which provides on 28 pages in three sections a total of 11 figures (S11-S21), three tables (S4-S6), and 19 equations (S1-S19) as additional material to the main text. Section S1 gives a detailed description of the GP6P parameterization, of the methods used for its evaluation and presents results from the parameterization as well as additional microscopic and macroscopic properties of GP6P. Section S2 proves that DFT/PMM-MFMD and DFT/PMM-MD yield very similar RDFs and that these RDFs do not depend on the chosen DFT setup. In Section S3 NMA's IR spectra calculated for all three DFT setups are presented in the frequency ranges [1800, 1000] cm^{-1} and [1380, 580] cm^{-1}.
[61]S. Bauer, P. Tavan, and G. Mathias, J. Chem. Phys. **140**, 104103 (2014).
[62]J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).
[63]J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 15 (2014).
[64]V. Kräutler, W. F. van Gunsteren, and P. Hünenberger, J. Comput. Chem. **22**, 501 (2001).
[65]H. C. Andersen, J. Chem. Phys. **52**, 24 (1983).
[66]W. C. Swope and H. C. Anderson, J. Chem. Phys. **76**, 637 (1982).
[67]M. P. Allen and D. Tildesley, in *Computer Simulations of Liquids* (Clarendon, Oxford, 1987), Chap. 2.4, pp. 46–50.
[68]D. C. Rapaport, in *The Art of Molecular Dynamics Simulation* (Cambridge University Press, Cambridge, UK, 2004), Chap. 2.3, pp. 18–20.
[69]M. P. Allen and D. Tildesley, in *Computer Simulations of Liquids* (Clarendon, Oxford, 1987), Chap. 1.5.2-3, pp. 24–29.
[70]G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[71]M. P. Allen and D. Tildesley, in *Computer Simulations of Liquids* (Clarendon, Oxford, 1987), Chap. 2.8, pp. 64–65.
[72]B. Schropp and P. Tavan, J. Phys. Chem. B **112**, 6233 (2008).
[73]B. Schropp and P. Tavan, J. Phys. Chem. B **114**, 2051 (2010).
[74]K. Ichikawa, Y. Kameda, T. Yamaguchi, H. Wakita, and M. Misawa, Mol. Phys. **73**, 79 (1991).
[75]W. E. Thiessen and A. H. Narten, J. Chem. Phys. **77**, 2656 (1982).
[76]S. Clough, Y. Beers, G. Klein, and L. Rothman, J. Chem. Phys. **59**, 2254 (1973).
[77]W. F. Murphy, J. Chem. Phys. **67**, 5877 (1977).
[78]C. Berweger, W. van Gunsteren, and F. Müller-Plathe, Chem. Phys. Lett. **232**, 429 (1995).
[79]R. A. Buckingham and J. Corner, Proc. R. Soc. London, Ser. A **189**, 118 (1947).

[80]G. S. Kell, J. Chem. Eng. Data **12**, 66 (1967).
[81]G. Jancso and W. A. Van Hook, Chem. Rev. **74**, 689 (1974).
[82]A. K. Soper, Chem. Phys. **258**, 121 (2000).
[83]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[84]G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[85]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[86]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[87]S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
[88]C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).
[89]M. Krack, Theor. Chem. Acc. **114**, 145 (2005).
[90]G. Lippert, J. Hutter, and M. Parrinello, Mol. Phys. **92**, 477 (1997).
[91]J. VandeVondele and J. Hutter, J. Chem. Phys. **118**, 4365 (2003).
[92]J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, J. Chem. Phys. **122**, 014515 (2005).
[93]J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy, J. Phys. Chem. B **113**, 11959 (2009).
[94]G. J. Martyna, M. L. Klein, and M. E. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).
[95]W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[96]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[97]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[98]T. Todorova, A. P. Seitsonen, J. Hutter, I.-F. W. Kuo, and C. J. Mundy, J. Phys. Chem. B **110**, 3685 (2006).
[99]N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[100]J. Perdew and W. Yue, Phys. Rev. B **33**, 8800 (1986).
[101]E. Vanden-Eijnden and G. Ciccotti, Chem. Phys. Lett. **429**, 310 (2006).
[102]M. Lingenheil, R. Denschlag, R. Reichold, and P. Tavan, J. Chem. Theory Comput. **4**, 1293 (2008).
[103]J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
[104]M. Martın, M. Aguilar, S. Chalmet, and M. Ruiz-López, Chem. Phys. **284**, 607 (2002).
[105]M. Sánchez, M. Aguilar, and F. O. del Valle, J. Comput. Chem. **18**, 313 (1997).
[106]M. Sánchez, M. Martın, M. Aguilar, and F. O. del Valle, J. Comput. Chem. **21**, 705 (2000).
[107]I. F. Galván, M. Sanchez, M. Martın, F. O. Del Valle, and M. Aguilar, Comput. Phys. Commun. **155**, 244 (2003).
[108]J. Borysow, M. Moraldi, and L. Frommhold, Mol. Phys. **56**, 913 (1985).
[109]The quality of various quantum correction factors[108] for the computation of IR spectra in the mid-IR region by FTTCF has been studied by Schmitz and Tavan[20] for a small organic molecule. Here, the so-called harmonic approximation correction factor was found to yield accurate results.
[110]G. Mathias and M. Baer, J. Chem. Theory Comput. **7**, 2028 (2011).
[111]G. Mathias, S. D. Ivanov, A. Witt, M. D. Baer, and D. Marx, J. Chem. Theory Comput. **8**, 224 (2011).
[112]J. A. Morrone and R. Car, Phys. Rev. Lett. **101**, 017801 (2008).
[113]D. C. Rapaport, in *The Art of Molecular Dynamics Simulation* (Cambridge University Press, Cambridge, UK, 2004), Chap. 4.2, pp. 84–89.
[114]R. C. Weast, *Handbook of Chemistry and Physics* (CRC, Boca Raton, 1983).
[115]K. Krynicki, C. Green, and D. Sawyer, Faraday Discuss. Chem. Soc. **66**, 199 (1978).
[116]K. Harris and L. Woolf, J. Chem. Eng. Data **49**, 1064 (2004).
[117]U. Kaatze, J. Chem. Eng. Data **34**, 371 (1989).
[118]R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[119]Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[120]T. Okabe, M. Kawata, Y. Okamoto, and M. Mikami, Chem. Phys. Lett. **335**, 435 (2001).
[121]P. Tröster and P. Tavan, J. Phys. Chem. Lett. **5**, 138 (2014).
[122]F. Siebert, Methods Enzymol. **246**, 501 (1995).
[123]Y. V. Venkatachalapathi, D. F. Mierke, J. P. Taulane, and M. Goodman, Biopolymers **26**, 763 (1987).
[124]L. C. Mayne and B. Hudson, J. Phys. Chem. **95**, 2962 (1991).
[125]A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[126]X. Chen, R. Schweitzer-Stenner, S. Krimm, N. G. Mirkin, and S. A. Asher, J. Am. Chem. Soc. **116**, 11141 (1994).
[127]P.-A. Cazade, F. Hédin, Z.-H. Xu, and M. Meuwly, J. Phys. Chem. B **119**, 3112 (2015).
[128]J.-J. Max and C. Chapados, J. Chem. Phys. **131**, 184505 (2009).
[129]Y. Tu and A. Laaksonen, J. Chem. Phys. **111**, 7519 (1999).
[130]M. Schwörer, C. Wichmann, E. Gawehn, and G. Mathias, J. Chem. Theory Comput. **12**, 992 (2016).

Der folgende Abdruck

Supplementary Information for
A Polarizable QM/MM Approach to the Molecular Dynamics of
Amide Groups Solvated in Water

Magnus Schwörer, Christoph Wichmann und Paul Tavan
J. Chem. Phys. **144**, 114504 (2016)

enthält zusätzliche Informationen zum oben abgedruckten Haupttext. Im Abschnitt S1 werden hier die im Haupttext nur grob skizzierte Konstruktion des PMM-Wassermodells GP6P sowie die Details der Evaluierung dokumentiert. Abschnitt S2 zeigt, dass der effiziente DFT/PMM mean-field-Ansatz sehr gute Vorhersagen für die radialen Verteilungsfunktionen in DFT/PMM-Systemen liefert, und dass diese unabhängig von der gewählten DFT-Methode sind. In Abschnitt S3 werden schließlich alle berechneten IR-Spektren im Spektralbereich von $1800\,\mathrm{cm}^{-1}$ bis $580\,\mathrm{cm}^{-1}$ zusammen mit GNC-Analysen gezeigt.

Supplementary Material to:

A Polarizable QM/MM Approach to the Molecular Dynamics of Amide Groups Solvated in Water

Magnus Schwörer,[a)] Christoph Wichmann[a)], and Paul Tavan[b)]
Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

[a)]These two authors contributed equally.
[b)]Electronic mail: paul.tavan@physik.uni-muenchen.de

S1. PARAMETERIZATION AND EVALUATION OF THE GP6P WATER MODEL

For the predominantly computational construction of our new Gaussian polarizable six-point (GP6P) potential, which should be particularly suited for hybrid DFT/PMM-MD simulations (cf. Section I), we adopted the general concepts underlying the parameterization[1,2] of the polarizable ν-point ($\nu = 4, 5, 6$) point charge water models called TLνP.[1] Because four- and five-point models, which feature three and four static partials charges, respectively, did not show[3] a density maximum near 277 K and because out-of-plane "lone-pair" charges are required for a proper modeling of the quadrupole moment, which steers the local order structures in the liquid phase,[4] we decided to choose a six-point geometry with five static Gaussian charge distributions for our new water model potential.

Figure 2 illustrates the geometry of the resulting GP6P model together with the distributions of the induced dipole, of the charges and of the mass-points. Procedural modifications of the TLνP parameterization strategy[1,2] became necessary, mainly because GP6P features with the widths σ_i of the Gaussian charge distributions at most three additional parameters (compared to TL6P).

A. Physical Corner Stones of GP6P

The construction of GP6P follows the general strategy[1,2] (i) to directly adopt experimentally well-measured properties of the water molecule in the gas and liquid phases wherever possible, (ii) to take advantage of the new DFT/PMM technology[5,6] (which includes the most recent version of the SAMM algorithm[7,8]) to compute most of the remaining electrostatic parameters, and (iii) to resort to empirical optimizations for as few parameters as possible.

1. Parameters Directly Adopted from Experiments

Like the point-charge TLνP potentials,[1,2] also our new Gaussian polarizable six-point model GP6P of H_2O is chosen rigid with the experimental liquid phase geometry[9,10] G_m, which is defined by the O-H distance $l_\mathrm{OH} = 0.968\,\text{Å}$ and the H-O-H angle $\varphi_\mathrm{HOH} = 105.3°$. Following the suggestion in Ref. 11 its distribution of static partial charges is constrained to yield the experimental gas phase dipole moment[12] $|\boldsymbol{\mu}^\mathrm{g}_\mathrm{exp}| = 1.855\,\mathrm{D}$. The electronic polarizability is represented by an induced Gaussian dipole distribution of width σ_μ, which is centered at the oxygen. The strength $\boldsymbol{\mu}^\mathrm{i}_\mathrm{PMM}$ of this induced dipole depends linearly and isotropically through the polarizability α_PMM, for which we choose[11] the experimental gas

phase value[13] $\alpha^{\mathrm{g}}_{\mathrm{exp}} = 1.47\,\text{Å}^3$, on the polarizing electric field, i.e. $\boldsymbol{\mu}^{\mathrm{i}}_{\mathrm{PMM}} = \alpha^{\mathrm{g}}_{\mathrm{exp}} \langle \mathbf{E}(\mathbf{r}_{\mathrm{O}}) \rangle_{\sigma_\mu}$. This field is an average over a Gaussian volume of width σ_μ centered at the position $\mathbf{r}_{\mathrm{O}}$ of the oxygen [cf. Eqs. (3) and (4) in Ref. 5].

2. Electrostatic Geometry of GP6P

The new DFT/PMM technology[5,6] is employed to compute almost all further features of the electrostatic signature of GP6P. According to Figure 2, this signature is characterized by two identical static Gaussian charge distributions of strengths $q_{\mathrm{H}} > 0$ and widths σ_{H} centered at the two hydrogen atoms. A negative static Gaussian charge distribution (strength q_{M}, width σ_{M}) sits at a massless site (M), which is located within the molecular triangle on the bisectrix of the bond angle φ_{HOH} at the distance l_{OM} from the oxygen. Furthermore, two massless so-called "lone-pair" sites (L) are found in the plane, which is defined by the bisectrix and the normal of the molecular plane. They are located symmetrically above and below that plane. The O-L distance is denoted as l_{OL}. The L-O-L angle φ_{LOL} is measured as indicated in Figure 2. The L-sites carry identical negative Gaussian charge distributions (strengths $q_{\mathrm{L}} < 0$, widths σ_{L}).

With the model's predefined vacuum dipole $\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{PMM}} = |\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{exp}}|\,\mathbf{e}_x$, where the unit vector $\mathbf{e}_x$ is oriented parallel to the H–O–H bisectrix in the O-M direction (cf. Figure 2), and with the charge neutrality $2q_{\mathrm{H}} + q_{\mathrm{M}} + 2q_{\mathrm{L}} = 0$ of the water molecule the electrostatic geometry G_{e} of GP6P is uniquely determined by the four parameters $\{l_{\mathrm{OM}}, l_{\mathrm{OL}}, \varphi_{\mathrm{LOL}}, q_{\mathrm{H}}\} \equiv G_{\mathrm{e}}$.

In the absence of a polarizing field, the electrostatic signature of the GP6P model is defined by G_{e} and by the three Gaussian widths σ_{H}, σ_{M} and σ_{L}. To reduce the number of parameters we choose $\sigma_{\mathrm{M}} = \sigma_{\mathrm{L}}$. and, thus, the remaining widths are determined by the ratio $\Sigma_{\mathrm{LH}} \equiv \sigma_{\mathrm{L}}/\sigma_{\mathrm{H}}$ and the value of σ_{H}. In the presence of a polarizing field also the width σ_μ of the induced Gaussian dipole at the oxygen contributes to the electrostatic signature.

As mentioned at the beginning of this section, almost all electrostatic parameters (i.e. G_{e}, Σ_{LH}, and σ_μ) will be determined computationally from DFT/PMM calculations. Here, only σ_{H} will be determined empirically from comparisons of PMM-MD simulations with experimental data on liquid water. This is a key difference to the optimization[1,2] of the TLνP point charge models, which determined *all* electrostatic parameters from DFT/PMM calculations.

3. Modeling the van der Waals Interactions

Beyond the Gaussian width σ_{H}, the empirical parameter optimization will also cover the parameters A_1, A_2, and B of the Buckingham[14] potential

$$U_{\mathrm{Bu}}(r \,|\, A_1, A_2, B) = A_1 \exp(-rA_2) - B/r^6, \tag{S1}$$

which is centered at the oxygen atom and models the van der Waals interactions between the GP6P water molecules at O-O distances r.

As targets of this empirical optimization, which will be carried out in the NVT_0 ensemble at the temperature $T_0 \equiv 300\,K$ and at the density[15] $n^{\mathrm{exp}}(p_0, T_0) = 0.9965$ g/cm^3 of liquid water at the standard pressure $p_0 \equiv 1$ atm, we chose the pressure p_0, and the experimental values for the mean potential energy $E_0^{\mathrm{exp}} = 9.92$ kcal/mol per molecule,[16] for the position[17] $r_{\mathrm{OO}}^{\mathrm{max,exp}} = 2.76\,$Å of the first peak of the O–O RDF $g_{\mathrm{OO}}(r)$, and for the thermal expansion coefficient[15] $\alpha_p^{\mathrm{exp}} = 2.8 \times 10^{-4}$/K.

To enable the later use of GP6P for solute-solvent systems, in which the van der Waals interactions of the solute atoms are modeled by LJ potentials, we will additionally determine from $U_{\mathrm{Bu}}(r \,|\, A_1, A_2, B)$ a related two-parameter LJ potential

$$U_{\mathrm{LJ}}(r \,|\, A, B) = A/r^{12} - B/r^6, \tag{S2}$$

which features the same dispersion parameter B as U_{Bu}. Its repulsive parameter A is obtained by fitting $U_{\mathrm{LJ}}(r \,|\, A, B)$ to $U_{\mathrm{Bu}}(r \,|\, A_1, A_2, B)$ in the range $r \in [2.4, 6]$ Å.[1]

B. Computational Methods

The iterative parameterization of GP6P followed the general scenario suggested in Ref. 1. It involved PMM-MD simulations and DFT/PMM calculations, which were carried out with the parallelized MD program package IPHIGENIE[18] and its parallelized DFT module CPMD.[19]

1. Setup of PMM-MD Simulations

All PMM-MD simulations, which were carried out during the iterative optimization[1] of the new GP6P model, employed the same (so-called medium size) simulation system, i.e. a periodic cubic box containing $N_{\mathrm{m}} = 1500$ water molecules. In $N_{\mathrm{m}}V_{\mathrm{m}}T$ simulations its volume V_{m} was generally (unless stated otherwise) chosen as $V_{\mathrm{m}} \equiv (35.58\,\text{Å})^3$ to yield the experimental[15] density $n^{\mathrm{exp}}(p_0, T_0)$ at our standard conditions. In such simulations the temperature T was controlled by a tightly coupled Bussi thermostat[20] (coupling time 0.1 ps).

In all Np_0T simulations the pressure was steered by a Berendsen barostat[21] characterized by a coupling time of 10 ps and by the isothermal compressibility $\kappa_T = 46 \times 10^{-6}$/atm, which is the experimental value[15] for liquid water at p_0 and T_0. All other simulation settings were chosen as described in Section II A of the main text.

As is also stated in Section II A, the pressure was calculated in all MD simulations from the virial expression.[22,23] This barometer was also applied during the parameterization of the predecessor models TL4P and TL5P.[1] In the course of the present study it turned out, however, that a simpler barometer expression,[23] which is solely applicable to additive force fields, was apparently employed during the parametrization[2] and sample application[3] of the TL6P model. Therefore, the TL6P model is flawed.

2. DFT Setup of Choice: MT/BP

Beyond PMM-MD simulations, the GP6P parameterization also covered DFT/PMM single-point calculations of selected water molecules embedded in solvation environments, which had been generated by PMM-MD. With the aim of staying compatible with previous DFT/(P)MM studies of the water molecule[1,11,24] we chose as DFT method the gradient-corrected exchange functional of Becke (B),[25] the correlation functional of Perdew (P),[26] combined with the norm-conserving pseudopotentials of Martins and Troullier (MT),[27] and a plane-wave cutoff of 80 Ry. This DFT method is denoted MT/BP. The selected DFT water molecule was always centered in a cubic box of volume $(9\,\text{Å})^3$ and the self-consistency convergence threshold[5] of the Kohn-Sham orbitals was set to $\chi_{\text{DFT}} = 10^{-7}$.

C. Procedures Executed at Each Parameterization Step

Each step $n = 1, 2, \ldots$ of the iterative parametrization comprised a set of sequentially executed procedures, which transformed the previous model GP6P^{n-1} into its successor GP6P^n. The iterations were repeated until a convergence of the model parameters was apparently reached.

1. Snapshot Set for DFT/PMM Single-Point Calculations

First a 400 ps PMM-MD simulation was performed in the $N_{\text{m}}V_{\text{m}}T_0$ ensemble for the model GP6P^{n-1} (GP6P$^0 \equiv$ TL6P). Five snapshots were drawn from the end of this trajectory at temporal distances of 10 ps. These snapshots are defined by the positions of all atoms and by the strengths $\boldsymbol{\mu}_{\text{PMM}}^{\text{i},k}$ of the induced dipoles of the PMM water models k. From each

snapshot, 300 water molecules were randomly selected as DFT fragments for subsequent DFT/PMM single-point calculations.

In this way we obtained, for each iteration step n, a supposedly representative set $\mathcal{S}^n$ of 1500 solvation structures s. The DFT/PMM calculations served to determine for each $s \in \mathcal{S}^n$ the dipole moment $\boldsymbol{\mu}_{\mathrm{DFT}}(s)$ and the charge density $\rho(s)$ of the DFT fragment, whose structure was kept fixed at the experimental liquid phase geometry G_{m}.

2. DFT/PMM Optimization of the Induced Dipole's Width σ_μ

For each solvation structure $s \in \mathcal{S}^n$, first the dipole moment

$$\boldsymbol{\mu}^{\mathrm{i}}_{\mathrm{DFT}}(s) = \boldsymbol{\mu}_{\mathrm{DFT}}(s) - |\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}|\,\mathbf{e}_x, \tag{S3}$$

which is induced in the DFT fragment by the PMM environment, is calculated. Here, $\boldsymbol{\mu}_{\mathrm{DFT}}(s)$ is the total dipole moment of the DFT fragment calculated by DFT/PMM and $|\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}| = 1.7920\,\mathrm{D}$ is the vacuum DFT dipole moment calculated[11] with MT/BP at G_{m}.

For comparison also the PMM analogues

$$\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s|\sigma_\mu) = \alpha^{\mathrm{g}}_{\mathrm{DFT}} \langle \mathbf{E}(\mathbf{r}_{\mathrm{O}}) \rangle_{\sigma_\mu} \tag{S4}$$

are computed for all $s \in \mathcal{S}^n$ by multiplying the local Gaussian average field $\langle \mathbf{E}(\mathbf{r}_{\mathrm{O}}) \rangle_{\sigma_\mu}$ introduced above in Section S1 A 1 with the DFT polarizability $\alpha^{\mathrm{g}}_{\mathrm{DFT}} = 1.6025\,\text{Å}^3$ obtained[11] by MT/BP for a water molecule, which was exposed to homogeneous external fields and kept fixed at G_{m}.[28]

The induced PMM dipoles $\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s|\sigma_\mu)$ depend on the Gaussian width σ_μ assumed for the local field average $\langle \mathbf{E}(\mathbf{r}_{\mathrm{O}}) \rangle_{\sigma_\mu}$. Therefore a variation of σ_μ can be employed to optimize the correlation between the induced DFT dipole moments $\boldsymbol{\mu}^{\mathrm{i}}_{\mathrm{DFT}}(s)$ and their PMM analogues $\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s|\sigma_\mu)$. For this purpose one can minimize the mean square deviation

$$\chi^2(\sigma_\mu) = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left[\mu^{\mathrm{i}}_{\mathrm{DFT},x}(s) - \hat{\mu}^{\mathrm{i}}_{\mathrm{PMM},x}(s \,|\, \sigma_\mu) \right]^2 \tag{S5}$$

between the x-components of the induced dipole moments $\boldsymbol{\mu}^{\mathrm{i}}_{\mathrm{DFT}}(s)$ obtained by DFT and their PMM analogues $\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s|\sigma_\mu)$ over the set $\mathcal{S}^n$ of solvation structures s.[1,24] Technically the minimization of $\chi^2(\sigma_\mu)$ was effected by scanning σ_μ from 0 to 1 Å in steps of 0.0005 Å. The resulting optimal width is denoted $\sigma^{\mathrm{opt}}_\mu$.

3. DFT/PMM Optimization of the Electrostatic Geometry

The next target of the parameter optimization is a DFT estimate $\hat{G}^n_{\mathrm{e}}$ of the electrostatic geometry.

First, for each solvation structure $s \in \mathcal{S}^n$ the electrostatic potential $\Phi_{\mathrm{DFT}}\left[\mathbf{r}_p|\rho(s)\right]$, which is generated by the charge density $\rho(s)$ of the DFT water molecule, is evaluated at the positions $\mathbf{r}_p$ of 500 points p, which are uniformly distributed on a sphere $\mathcal{P}$ of radius $R = 2.76\,\text{Å}$ surrounding the molecule's center of geometry. These points are given in a frame of reference, which is fixed with respect to the molecular geometry. The radius of the sphere coincides with the location of the first maximum of the oxygen-oxygen radial distribution function (RDF) of water at standard conditions.[17]

Next, the DFT fragment is replaced by a PMM water model, whose Gaussian induced dipole $\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s\,|\,\sigma_\mu^{\mathrm{opt}})$ [Eq. S4] has the above determined width $\sigma_\mu^{\mathrm{opt}}$ and is located at $\mathbf{r}_{\mathrm{O}}$. The potential $\Phi_{\boldsymbol{\mu}}\left[\mathbf{r}_p\,|\,\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s\,|\,\sigma_\mu^{\mathrm{opt}})\right]$ generated by $\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s\,|\,\sigma_\mu^{\mathrm{opt}})$ is then calculated at all points $\mathbf{r}_p$ of the spherical surface.

For each solvation structure s these two surface potentials are subtracted at all points $\mathbf{r}_p$ yielding the difference potentials

$$\Delta\Phi_{\mathrm{DFT}}\left(\mathbf{r}_p|s\right) \equiv \Phi_{\mathrm{DFT}}\left[\mathbf{r}_p|\rho(s)\right] - \Phi_{\boldsymbol{\mu}}\left[\mathbf{r}_p\,|\,\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s\,|\,\sigma_\mu^{\mathrm{opt}})\right]. \tag{S6}$$

If one assumes that the fluctuations of the electrostatic field, which acts on the selected DFT water molecule in a liquid phase environment, mainly change its induced dipole moment, which should be very well approximated by its PMM analogue, then the values of the difference potential $\Delta\Phi_{\mathrm{DFT}}\left(\mathbf{r}_p|s\right)$ should be similar in all solvation structures $s \in \mathcal{S}^n$. Thus, the difference potential should only slightly fluctuate[1] around the static contribution $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p)$ to the surface potential $\Phi_{\mathrm{DFT}}\left[\mathbf{r}_p|\rho(s)\right]$, which is generated by the average higher moments of the DFT charge density.

Averaging $\Delta\Phi_{\mathrm{DFT}}\left(\mathbf{r}_p|s\right)$ over all solvation structures yields the mean difference potential

$$\langle\Delta\Phi_{\mathrm{DFT}}(\mathbf{r}_p)\rangle_{\mathcal{S}} \equiv \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\Delta\Phi_{\mathrm{DFT}}\left(\mathbf{r}_p|s\right), \tag{S7}$$

at all points $p \in \mathcal{P}$, whose surface average variance

$$\sigma^2_{\mathcal{S}} \equiv \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}}\left\{\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\left[\Delta\Phi_{\mathrm{DFT}}\left(\mathbf{r}_p|s\right) - \langle\Delta\Phi_{\mathrm{DFT}}(\mathbf{r}_p)\rangle_{\mathcal{S}}\right]^2\right\} \tag{S8}$$

should be small. Denoting the mean difference potential as

$$\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p) \equiv \langle\Delta\Phi_{\mathrm{DFT}}(\mathbf{r}_p)\rangle_{\mathcal{S}}, \tag{S9}$$

the variation of this static part of the potential $\Phi_{\mathrm{DFT}}\left[\mathbf{r}_p|\rho(s)\right]$ on the spherical surface $\mathcal{P}$ is

$$\sigma^2_{\mathcal{P}} \equiv \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}}\left[\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p) - \langle\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p)\rangle_{\mathcal{P}}\right]^2. \tag{S10}$$

Then the ratio $\sigma_{\mathcal{S}}/\sigma_{\mathcal{P}}$ measures the size of the fluctuations of the difference potential $\Delta\Phi_{\mathrm{DFT}}(\mathbf{r}_p|s)$ in the snapshot set $\mathcal{S}^n$ around its mean $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p)$ on the scale $\sigma_{\mathcal{P}}$, which is defined by the variations of $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p)$ on the spherical surface. According to the above conjecture, this ratio should be small.

The DFT estimate $\hat{G}^n_{\mathrm{e}} = \{l_{\mathrm{OM}}, l_{\mathrm{OL}}, \varphi_{\mathrm{LOL}}, \hat{q}_{\mathrm{H}}\}^n$ of the electrostatic geometry G_{e}, which approximates the average higher moments of the DFT fragment (at the predefined value $|\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}|$ of the static dipole moment) within the chosen model class as accurately as possible, is finally found by minimizing over all surface points the functional

$$\xi^2(\hat{G}_{\mathrm{e}}) = \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}}\left[\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p) - \Phi^{\mathrm{stat}}_{\mathrm{PMM}}(\mathbf{r}_p|\hat{G}_{\mathrm{e}})\right]^2 \tag{S11}$$

of the four-parameter electrostatic geometry $\hat{G}_{\mathrm{e}}$. This functional is the mean square deviation between the static contribution $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p)$ [Eq. (S9)] to the DFT potential $\Phi_{\mathrm{DFT}}[\mathbf{r}_p|\rho(s)]$ and the potential $\Phi^{\mathrm{stat}}_{\mathrm{PMM}}(\mathbf{r}_p|\hat{G}_{\mathrm{e}})$, which is generated by the static Gaussian charges of an unpolarized ($\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}} = \mathbf{0}$) PMM water model with the electrostatic geometry $\hat{G}_{\mathrm{e}}$. These charges have the Gaussian widths $\sigma^{n-1}_{\mathrm{H}}$ and $\sigma^{n-1}_{\mathrm{L}}$ of the model GP6P^{n-1}. The four-parameter space spanned by $\hat{G}_{\mathrm{e}}$ is searched by the Levenberg-Marquardt algorithm,[29,30] which yields robust results.

4. DFT/PMM Computation of the Gaussian Width Ratio Σ_{LH}

At the thus obtained electrostatic geometry $\hat{G}^n_{\mathrm{e}}$, first estimates of the Gaussian charge widths $\hat{\sigma}^n_{\mathrm{H}}$ and $\hat{\sigma}^n_{\mathrm{L}}$ are computed by minimizing the functional

$$\xi'^2(\hat{\sigma}_{\mathrm{H}}, \hat{\sigma}_{\mathrm{L}}) = \frac{1}{|\mathcal{P}'|}\sum_{p\in\mathcal{P}'}\left[\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r}_p) - \Phi^{\mathrm{stat}}_{\mathrm{PMM}}(\mathbf{r}_p|\hat{G}^n_{\mathrm{e}}, \hat{\sigma}_{\mathrm{H}}, \hat{\sigma}_{\mathrm{L}})\right]^2 \tag{S12}$$

with respect to the widths. This functional is analogous to Eq. (S11) but is based on a smaller sphere $\mathcal{P}'$ of radius $R' = 1.70\,\text{Å}$, which roughly corresponds to the distance, at which the first peak of the oxygen-hydrogen RDF is found in liquid water at standard conditions.[17] Because the thus obtained Gaussian widths $\hat{\sigma}^n_{\mathrm{L}}$ and $\hat{\sigma}^n_{\mathrm{H}}$ will turn out to be smaller than about $0.7\,\text{Å}$, the potential at the surface of the large sphere $\mathcal{P}$ is quite insensitive to changes of $\hat{\sigma}_L$ and $\hat{\sigma}_H$, whereas the surface potential of the small sphere $\mathcal{P}'$ is much more sensitive.

Consequently, the computation of the widths by minimizing $\xi'^2(\hat{\sigma}_{\mathrm{H}}, \hat{\sigma}_{\mathrm{L}})$ mainly serves to bring the radial decay of the PMM model potential $\Phi^{\mathrm{stat}}_{\mathrm{PMM}}(\mathbf{r})$ close to that of its DFT antetype $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}(\mathbf{r})$ by optimizing their mean square difference on two concentric spherical surfaces $\mathcal{P}'$ and $\mathcal{P}$, which surround the center of geometry of the water molecule with the radii $1.70\,\text{Å}$ and $2.76\,\text{Å}$, respectively. The resulting widths $\hat{\sigma}_{\mathrm{H}}$ and $\hat{\sigma}_{\mathrm{L}}$ eventually define the

ratio $\Sigma^n_{\mathrm{LH}} = \hat{\sigma}^n_{\mathrm{L}}/\hat{\sigma}^n_{\mathrm{H}}$, which is kept constant during the further refinement in the current iteration n. Thus, only one Gaussian width (σ_{H}) of a charge distribution remains as a free parameter to be computed by empirical optimization.

5. *Transforming the DFT- into an Experiment-Adapted Model*

DFT descriptions of molecules generally differ from their real antetypes. Because the desired PMM model should approximate a real water molecule, but not its DFT description, as accurately as possible, the above DFT-based computation of the electrostatic geometry $\hat{G}_{\mathrm{e}}$ requires a corresponding correction. For this purpose the strengths $\{\hat{q}_{\mathrm{H}}, \hat{q}_{\mathrm{M}}, \hat{q}_{\mathrm{L}}\}$ of the static PMM partial charge distributions are scaled by the factor

$$\gamma \equiv |\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{exp}}|/|\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}| = 1.03515625 \tag{S13}$$

to yield the experiment-adapted parameter set $\{q_{\mathrm{H}}, q_{\mathrm{M}}, q_{\mathrm{L}}\}$. This scaling transforms the DFT estimate $\hat{G}_{\mathrm{e}}$ featuring the static DFT dipole moment[11] $|\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}| = 1.7920\,\mathrm{D}$ into the electrostatic geometry $G_{\mathrm{e}} = \{l_{\mathrm{OM}}, l_{\mathrm{OL}}, \varphi_{\mathrm{LOL}}, q_{\mathrm{H}}\}$ of an "experimental" PMM model, which then has the static dipole moment $|\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{PMM}}| = |\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{exp}}| = 1.855\,\mathrm{D}$.

Furthermore, the polarizability of the PMM model, for which the DFT value[11] $\alpha^{\mathrm{g}}_{\mathrm{DFT}} = 1.6025\,\text{Å}^3$ has been assumed in Section S1 C 2 in the computation of the induced dipole's width σ^n_μ, is now set to the experimental value,[13] i.e. $\alpha_{\mathrm{PMM}} = \alpha^{\mathrm{g}}_{\mathrm{exp}} = 1.47\,\text{Å}^3$.

6. *Empirical PMM-MD Optimization of σ_{H}, A_1, and A_2*

The Gaussian width σ_{H} and, through the constant Gaussian width ratio Σ^n_{LH} (c.f. Section S1 C 4), also the width $\sigma_{\mathrm{L}} = \sigma_{\mathrm{H}}\Sigma^n_{\mathrm{LH}}$ are next refined in an empirical optimization by a subsequent weak-coupling (WC) PMM-MD simulation.[31] This simulation additionally serves to determine the parameters A_1 and A_2 of the Buckingham potential U_{Bu}, whereas its dispersion parameter B is kept fixed at the value B^n of the previous iteration. The initial value $B^0 \equiv 663\,\text{Å}^6\,\mathrm{kcal/mol}$ is adopted from the TL6P model.[2] Thus, the three parameters σ_{H}, A_1, and A_2 are optimized by WC PMM-MD simulations in the $N_{\mathrm{m}}V_{\mathrm{m}}T_0$ ensemble at our standard conditions employing an integration time step of $\Delta t = 1$ fs.

As we have indicated in Section S1 A 3, the targets of the WC simulations are the mean potential energy E^{exp}_0 per molecule, the pressure p_0, and the position $r^{\mathrm{max,exp}}_{\mathrm{OO}}$ of the first peak of the RDF $g_{\mathrm{OO}}(r)$. Here, σ_{H} is negatively coupled to $E - E^{\mathrm{exp}}_0$ with the effective coupling time[31] $\tau_\sigma/C_\sigma = 20\,\mathrm{ps\,kcal/(mol\,Å)}$, i.e.

$$\sigma_{\mathrm{H}}(t + \Delta t) = \sigma_{\mathrm{H}}(t) - (C_\sigma/\tau_\sigma)[E(t) - E^{\mathrm{exp}}_0]\Delta t \tag{S14}$$

(t denotes the current time step) and A_2 is coupled to p with $\tau_p/C_p = 0.1\,\mathrm{ps\,Å\,atm}$, i.e.

$$A_2(t+\Delta t) = A_2(t) + (C_p/\tau_p)[p(t) - p_0]\Delta t. \tag{S15}$$

A_1 is indirectly coupled to the position[17] $r_{\mathrm{OO}}^{\mathrm{max,exp}} = 2.76\,\mathrm{Å}$ of the first peak of $g_{\mathrm{OO}}(r)$ by negatively coupling the value

$$U'(t) \equiv U_{\mathrm{Bu}}(r_{\mathrm{a}}|\, A_1(t), A_2(t), B^n) \tag{S16}$$

of the current Buckingham potential at $r_{\mathrm{a}} = 2.3$ Å to $r_{\mathrm{OO}}^{\mathrm{max,exp}} = 2.76\,\mathrm{Å}$. As effective coupling time for $U'(t)$ we chose $\tau_{U'}/C_{U'} = 0.125\,\mathrm{ps\,Å\,/(kcal/mol)}$. An estimate of the associated RDF $g_{\mathrm{OO}}(r)$ and thus of $r_{\mathrm{OO}}^{\mathrm{max}}$ is recalculated every 250 fs from the last $\tau_{\mathrm{a}} \equiv 250\,\mathrm{fs}$ of WC PMM-MD. Denoting the time of this recalculation by $t_n \equiv n\tau_{\mathrm{a}}$ with $n = 1, 2, \ldots$ the difference $\Delta r^{\mathrm{max}}(t_n) \equiv r_{\mathrm{OO}}^{\mathrm{max}}(t_n) - r_{\mathrm{OO}}^{\mathrm{max}}$ yields the target value

$$U'(t_{n+1}) = U'(t_n) - (C_{U'}/\tau_{U'})\Delta r_{\mathrm{OO}}^{\mathrm{max}}(t_n)\tau_{\mathrm{a}}. \tag{S17}$$

A linear interpolation between the values of U', which are defined with respect to the coarse time scale τ_{a}, yields after a few lines of algebra the WC dynamics

$$U'(t+\Delta t) = U'(t) - (C_{U'}/\tau_{U'})\Delta r_{\mathrm{OO}}^{\mathrm{max}}(t_n)\Delta t \tag{S18}$$

for each time point t with $t_n < t \leq t_{n+1}$. Thus one gets at each time point t with Eqs. (S16) and (S1) the prediction

$$A_1(t+\Delta t) = \left\{U'(t+\Delta t) + B^n/r^6\right\} \exp[A_2(t+\Delta t)r_{\mathrm{a}}] \tag{S19}$$

of the last missing parameter. After 400 ps of WC simulation, temporal averages $\langle\sigma_{\mathrm{H}}(t)\rangle_\tau$, $\langle A_1(t)\rangle_\tau$, and $\langle A_2(t)\rangle_\tau$, are collected for the next $\tau \equiv 200\,\mathrm{ps}$ of WCMD, which are subsequently identified with the parameters of the model GP6P^n. For the very last WC simulation, which eventually yielded the final model parameters, all WC coupling constants were increased by a factor of 5.

7. Tuning of the Dispersion Parameter B

The empirical optimization of the dispersion parameter B requires a further target observable. We chose the thermal expansion coefficient α_p, which has at our standard conditions the experimental value[15] $\alpha_p^{\mathrm{exp}} = 2.8 \times 10^{-4}/\mathrm{K}$. For its optimal choice we conduct a series of further WC simulations (as described just above) targeting σ_{H}, A_1, and A_2, for 12 B values in the range $[399, 1125]\,\mathrm{Å}^6$ kcal/mol, which were chosen as $B = B^0 + i\Delta B$ with $\Delta B = 66\,\mathrm{Å}^6$ kcal/mol and $i \in \{-4, -3, \ldots, 7\}$. For each of these 12 models, the associated thermal

expansion coefficient $\alpha_p(B)$ is computed following Section S1 D below. The value for B^n is selected by linear interpolation of $\alpha_p(B)$ to the targeted experimental value α_p^{exp}. With this value B^n another WC simulation yields the final model parameters σ_{H}, A_1, and A_2 of GP6P^n. This model defines the initial conditions of the next cycle.

8. A Shortcut of the Optimization Cycle

Obviously the determination of B^n described above is computationally quite demanding. Therefore, the dispersion parameter B was kept at its initial value B^0 during the first four optimization cycles, which solely comprised all steps described in the paragraphs S1 C 1-S1 C 6. At the end of the fourth cycle B was optimized as described in Section S1 C 7 yielding B^4 and a two more optimization cycles (excluding a further optimization of B) served to check whether the remaining parameters showed an apparent convergence.

D. Methods for the Evaluation of GP6P

For a characterization of the GP6P model of liquid water at p_0 and T_0 we almost exactly copied the computational setups explained in Section 3.2 of Ref. 1. Here, the only difference is that we generally chose the simulation box of intermediate size containing $N_{\mathrm{m}} = 1500$ water molecules instead of a smaller box containing only $N_{\mathrm{s}} = 728$ molecules. In some cases, in which the size dependence of calculated observables had to be studied, we also considered a large box with $N_{\mathrm{l}} = 3374$ GP6P models. The general PMM-MD simulation settings, including the thermostat and barostat settings, were chosen as described in Sections II A and S1 B 1.

To asses the targeted properties of GP6P at our standard conditions ($p_0 = 1$ atm, $T_0 = 300\,\mathrm{K}$) we calculated the potential energy E_0 per molecule, the pressure p, and the RDF $g_{\mathrm{OO}}(r)$ as averages from a 3 ns MD trajectory simulated in the $N_{\mathrm{m}}V_{\mathrm{m}}T_0$ ensemble. The density $n(p_0, T_0)$ was averaged over a 3 ns $N_{\mathrm{m}}p_0T_0$ simulation.

The thermal expansion coefficient α_p, which had been the fourth target of the parametrization, was determined from two 3 ns $N_{\mathrm{m}}p_0T_\pm$ simulations, with temperatures $T_\pm = T_0 \pm 10\,\mathrm{K}$. Numerical differentation of the logarithm of the density with respect to the temperature according to Eq. (13) in Ref. 1 then yielded α_p. From the same trajectories, also GP6P predictions for the heat capacity C_p at constant pressure was extracted by numerical differentiation of the average total energy per molecule with respect to the temperature using Eq. (12) in Ref. 1.

The isothermal compressibility κ_T was obtained from two 3 ns $N_{\mathrm{m}}V_\pm T_0$ simulations, in which the volumes $V_\pm$ were chosen such that the densities were $n_+ = 1.047\ \mathrm{g/cm^3}$ or $n_- =$

0.947 g/cm^3, by numerical differentation of the logarithm of the density with respect to the average pressure according to Eq. (11) in Ref. 1.

Applying Eq. (9) in Ref. 1, finite size self-diffusion constants $D_0(N_i)$ were calculated for all three box sizes V_i ($i \in \{\mathrm{s}, \mathrm{m}, \mathrm{l}\}$) from $N_i V_i T_0$ simulations at $n^{\exp}(p_0, T_0)$ using a weak-coupling Berendsen[32] thermostat with the very large coupling time of 5 ps for temperature control. At such a large coupling time the perturbation of the dynamics by this thermostat becomes negligible.[33] One has, however, to choose target temperatures $\tilde{T}_{\mathrm{Be}}(N_i) < T_0$, which guarantee the average simulation temperatures $\langle T \rangle$ are close to T_0. For each of the three box sizes N_i, the values $D_0(N_i)$ were calculated from ten independent $N_i V_i T_0$ simulations of 250 ps duration.

Because D_0 strongly increases with the temperature,[34] we corrected the errors introduced by the deviations of the actual average temperature $\langle T \rangle$ from the target temperature T_0. For this purpose, we interpolated the experimental[34] function $D_0^{\exp}(T)$ in the temperature range $T \in [283.2, 323.2]\,\mathrm{K}$ by a second-order polynomial $f_2(T)$. Assuming that the temperature-dependence of $D_0(N_i, \langle T \rangle)$ is well approximated by $D_0^{\exp}(T)$, the corrected quantity $\tilde{D}_0(N_i, T_0) \equiv [f_2(T_0)/f_2(\langle T \rangle)] D_0(N_i, \langle T \rangle)$ was obtained by scaling. The self-diffusion constant D_0 in the infinite system and the viscosity η were then determined by extrapolation from the finite size values $\tilde{D}_0(N_i, T_0)$ employing Eq. (10) in Ref. 1.

Also the dielectric constant ϵ of the infinite system was derived by extrapolation from the finite size values $\epsilon(N_i)$, which were self-consistently obtained from the fluctuations of the total dipole moment through the corrected version[35] of Eq. (8) in Ref. 1 using perturbation theory.[36] Here, for each box size N_i a total of 10 zero'th order $N_i V_i T_0$ simulations were executed for 5 ns assuming a dielectric constant $\epsilon_{\mathrm{RF}} = 78$ for the surrounding dielectric continuum.[37]

The only observable, which we computed as a check on the transferability of the GP6P model to other thermodynamic conditions than the one (p_0,T_0) used during the parameterization, was the temperature-density profile $n(p_0, T)$ in the temperature range $T \in [250, 320]\,\mathrm{K}$. For its determination we carried a replica exchange simulation[38–40] in the generalized $N_{\mathrm{m}} p_0 \tilde{T}_k$ ensemble, with a temperature ladder comprising 15 rungs $k = 0, \ldots, 14$ with temperatures $\tilde{T}_k = (250 + 5k)\,\mathrm{K}$. Employing the deterministic even-odd scheme[41] temperature exchanges were attempted every 10 ps. The total simulation time was 20 ns per rung.

E. Results of the GP6P Parametrization

During the iterative parameterization described in Section S1 C several target quantities were optimized. For a few of these quantities we now document the final stage of the optimization. Here we focus on the observables $\chi^2(\sigma_\mu)$ [Eq. (S5)] and $\alpha_p(B)$ (cf. Section

S1 C 7), which were employed to determine the Gaussian width σ_μ of the induced PMM dipole $\boldsymbol{\mu}^{\mathrm{i}}_{\mathrm{PMM}}$ and the dispersion attraction parameter B entering the Buckingham potential $U_{\mathrm{Bu}}(r)$ [Eq. (S1)], respectively.

1. Dipole Correlations and Distributions

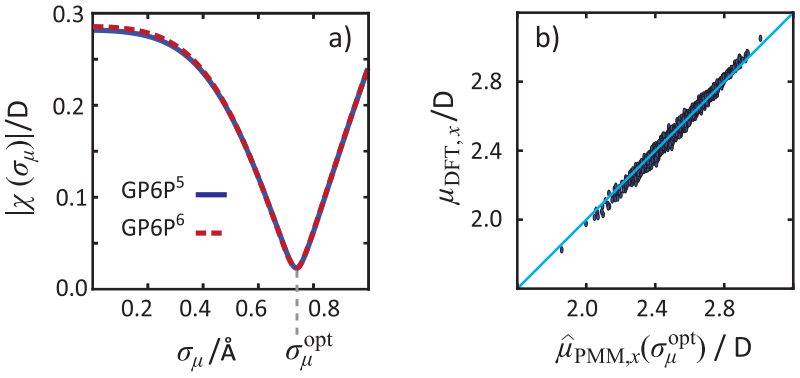


FIG. S11. a) Root mean square deviation $|\chi(\sigma_\mu)|$ of the induced Gaussian PMM dipoles from their DFT antetypes as a function of the Gaussian width σ_μ for the snapshot sets $\mathcal{S}^n$ in the steps $n = 5$ (blue) and $n = 6$ (red) of the iterative parametrization. b) Correlations between the x-components of the total dipole moments $\mu_{\mathrm{DFT},x}$ and $\hat{\mu}_{\mathrm{PMM},x}(\sigma_\mu^{\mathrm{opt}})$ for the snapshot set $\mathcal{S}^6$. As a guide to the eye the light blue line shows the hypothetical exact correlation.

Figure S11(a) shows the root of the mean square deviation $\chi^2(\sigma_\mu)$ defined by Eq. (S5) as a function of the PMM dipole's Gaussian width σ_μ. The observable $\chi^2(\sigma_\mu)$ measures the quality, at which the PMM dipoles $\hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}$ determined by the linear response Eq. (S4) reproduce their DFT/PMM counterparts $\boldsymbol{\mu}^{\mathrm{i}}_{\mathrm{DFT}}$ defined by Eq. (S3). Its minimum marks the optimal choice $\sigma_\mu^{\mathrm{opt}}$ for which the fifth and sixth iteration both yield the value 0.739 Å indicating that the optimization of σ_μ is converged.

Figure S11(b) depicts the correlations between the x-components of the total dipole moments

$$\boldsymbol{\mu}_{\mathrm{DFT}} = \boldsymbol{\mu}^{\mathrm{i}}_{\mathrm{DFT}}(s) + \boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}$$

of the DFT fragment and of the total dipoles moments

$$\hat{\boldsymbol{\mu}}_{\mathrm{PMM}}(\sigma_\mu^{\mathrm{opt}}) = \hat{\boldsymbol{\mu}}^{\mathrm{i}}_{\mathrm{PMM}}(s|\sigma_\mu^{\mathrm{opt}}) + \boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}$$

of the DFT-adjusted PMM model for the solvation structures $s \in \mathcal{S}^6$. Here, the induced PMM dipoles were computed by linear response using the optimal Gaussian width $\sigma_\mu^{\mathrm{opt}}$ and

the DFT vacuum polarizability $\alpha^{\mathrm{g}}_{\mathrm{DFT}}$. Furthermore recall that the DFT-adjusted PMM model is defined by the rigid electrostatic geometry $\hat{G}_{\mathrm{e}}$, which generates the DFT vacuum dipole moment $\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}$ (cf. also Section S1 C 5). Figure S11(b) demonstrates that the DFT dipole moments $\boldsymbol{\mu}_{\mathrm{DFT}}$ correlate very well with their PMM models $\hat{\boldsymbol{\mu}}_{\mathrm{PMM}}$ in the PMM snapshot set $\mathcal{S}^6$.

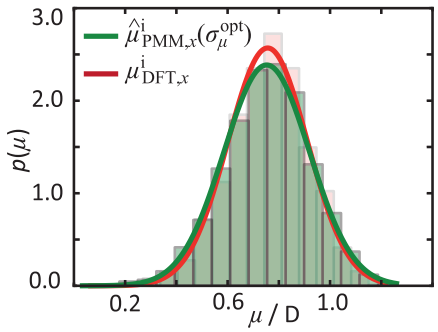


FIG. S12. Normalized histograms $p(\mu)$ and associated normal distributions (lines) for the induced dipole moments $\mu^{\mathrm{i}}_{\mathrm{DFT},x}$ of the respective DFT fragments (red) and $\hat{\mu}^{\mathrm{i}}_{\mathrm{PMM},x}$ of their DFT-adjusted PMM models (green) in the snapshot set $\mathcal{S}^6$.

Figure S12 provides an alternative statistical representation of the data shown already in Figure S11(b). The figure compares for the respective x-components the distributions of the induced dipole moments calculated for the 1500 solvation structures $s \in \mathcal{S}^6$ by DFT/PMM and by PMM. The averages $\langle \mu^{\mathrm{i}}_{\mathrm{DFT},x} \rangle_{\mathcal{S}} = 0.756\,\mathrm{D}$ and $\langle \hat{\mu}^{\mathrm{i}}_{\mathrm{PMM},x} \rangle_{\mathcal{S}} = 0.753\,\mathrm{D}$ are almost equal. The standard deviation $\sigma(\mu^{\mathrm{i}}_{\mathrm{DFT},x}) = 0.155\,\mathrm{D}$ is a little smaller than $\sigma(\hat{\mu}^{\mathrm{i}}_{\mathrm{PMM},x}) = 0.167\,\mathrm{D}$. Thus, the fluctuations of the induced DFT dipole moments around their average value amount within the set $\mathcal{S}^6$ to 20 %. The similarity of the two distributions in the snapshot set $\mathcal{S}^6$ confirms once again that the induced Gaussian PMM dipoles approximate their DFT antetypes very well.

2. $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}$ *is Actually Almost Stationary*

The optimization of the rigid model $\hat{G}_{\mathrm{e}}$ for the electrostatic geometry of the GP6P model and of the Gaussian width ratio Σ_{LH} [cf. Eqs. (S11) and (S12)], rests on the assumption that the potential $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}$ defined by Eq. (S9) is essentially stationary on the spherical surfaces $\mathcal{P}$ and $\mathcal{P}'$ in a snapshot set $\mathcal{S}^n$. In Section S1 C 3 we have introduced the ratio $\sigma_{\mathcal{S}}/\sigma_{\mathcal{P}}$ as a stationarity measure [this ratio is determined by Eqs. (S8) and (S10)]. Here, small values indicate stationarity.

For the large spherical surface surrounding the selected water molecule we found a ratio of 3.5 % i.e. the fluctuations $\sigma_{\mathcal{S}}$ of $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}$ on the surface $\mathcal{P}$ amount only to this small percentage

of its total variation $\sigma_{\mathcal{P}}$ on $\mathcal{P}$. On the smaller sphere $\mathcal{P}'$ the corresponding ratio is 4.0 % and, thus, is also quite small. As a result, the DFT/PMM calculations demonstrate that the stationarity assumption underlying the optimization of $\hat{G}_{\mathrm{e}}$ and Σ_{LH} is valid to a very good approximation.

3. Radial Decay of the Electrostatic Potential: PMM vs. DFT

The electrostatic geometry $\hat{G}_{\mathrm{e}}$ and the ratio $\Sigma_{\mathrm{LH}} = \sigma_{\mathrm{L}}/\sigma_{\mathrm{H}}$ of the Gaussian widths of the negative (σ_{L}) and positive (σ_{H}) charge distributions were optimized by considering the electrostatic potential on the surfaces of the two differently sized spheres $\mathcal{P}$ and $\mathcal{P}'$ (see Sections S1 C 3 and S1 C 4). The use of two different radii was supposed to guarantee that the radial decay of $\Phi^{\mathrm{stat}}_{\mathrm{DFT}}$ and of its model $\Phi^{\mathrm{stat}}_{\mathrm{PMM}}$ resemble each other at short distances from the center of geometry of a water molecule in a PMM environment.

For a verification of this resemblance it suffices to consider the *total* electrostatic potentials Φ_{DFT} and Φ_{PMM} generated by the DFT fragment and its DFT-adjusted PMM model, because these potentials both differ from the associated *stationary* potential by the potential $\Phi_{\boldsymbol{\mu}}$, which is caused by the induced Gaussian PMM dipole. Furthermore, this resemblance should become apparent, if one considers the radial variations of the potentials along lines directed from the O atom to the charges at the H and L sites. We denote the associated radial distances as r_{OH} and r_{OL}, respectively. In the liquid phase these directions are important, because they roughly point toward the nearest atoms of hydrogen bonded neighbors. Values $r_{\mathrm{OH}} \in [2.2, 3.2]$ Å and $r_{\mathrm{OL}} \in [1.3, 2.3]$ Å cover the first peaks of the O-O and O-H RDFs.[17]

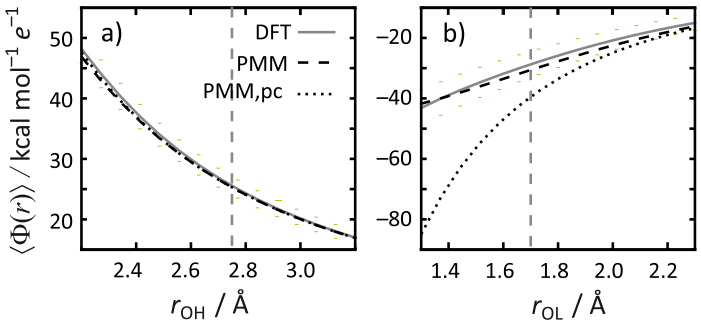


FIG. S13. Radial changes of the average electrostatic potentials $\langle\Phi(r)\rangle_{\mathcal{S}}$ along the coordinates r_{OH} (a) and r_{OL} (b) in the snapshot set $\mathcal{S}^6$. $\langle\Phi_{\mathrm{DFT}}(r)\rangle_{\mathcal{S}}$ (gray solid) is compared to $\langle\Phi_{\mathrm{PMM}}(r)\rangle_{\mathcal{S}}$ (black dashed) and the potential $\langle\Phi^{\mathrm{pc}}_{\mathrm{PMM}}(r)\rangle_{\mathcal{S}}$ of an associated point charge PMM model (black dotted).

Figure S13 shows for the two coordinates r_{OH} (a) and r_{OL} (b) the radial decay characteristics of the average electrostatic potentials $\langle\Phi(r)\rangle_{\mathcal{S}}$, which are generated either by the

DFT fragments (gray solid), or by DFT-adjusted PMM models replacing these fragments (black dashed), and by equivalent PMM models with point-like partial charges (black dotted). In the liquid phase one expects[17] the nearest neighbor molecules at $r_{\mathrm{OH}} \approx 2.8\,\text{Å}$ and $r_{\mathrm{OL}} \approx 1.7\,\text{Å}$, respectively. These distances are marked by vertical gray dashed lines.

The variations with r_{OH} are fairly similar for all three potentials depicted in Figure S13(a). In contrast, such a similarity is found in Figure S13(b) solely at large distances $r_{\mathrm{OL}} \gtrsim 2\,\text{Å}$, whereas at distances $r_{\mathrm{OL}} \lesssim 2\,\text{Å}$, the point charge potential $\langle \Phi^{\mathrm{pc}}_{\mathrm{PMM}}(r_{\mathrm{OL}}) \rangle_{\mathcal{S}}$ is much steeper than the other two potentials, which are quite similar also here. Hence, partial point charges exert much stronger forces on neighboring molecules than Gaussian charge models or the quantum mechanical charge distributions of DFT fragments. As a result, for a close modeling of quantum mechanical antetypes the use of Gaussian partial charges seems to be mandatory.

4. Determining the Dispersion Parameter B

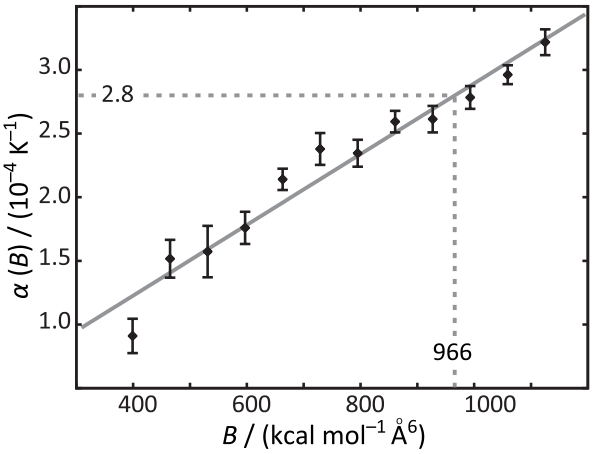


FIG. S14. The thermal expansion coefficient $\alpha_p(B)$ calculated for GP6P[4] models with different values for the dispersion parameter B. and the linear interpolation (solid gray) which predicts $B = 966\,\text{Å}^6$ kcal/mol to yield the target experimental value[15] $\alpha_p = 2.8 \times 10^{-4}/\text{K}$ (horizontal line).

As is explained in Section S1 C 7, the van der Waals dispersion parameter B [cf. Eqs. (S1) and (S2)] of GP6P was empirically determined by targeting the experimental value[15] $\alpha_p^{\mathrm{exp}} = 2.8 \times 10^{-4}/\text{K}$ of the isobaric thermal expansion coefficient α_p. For this purpose, during step $n = 4$ of the iterative optimization, 12 GP6P[4] models with different B parameters in the range $[399, 1125]\,\text{Å}^6$ kcal/mol were constructed by WC simulations (i.e. they differ also in A_1, A_2, σ_{H}, and σ_{L}), and $\alpha_p(B)$ was calculated by the methods described in Section S1 D.

Figure S14 shows $\alpha_p(B)$ (dots) together with error bars calculated by block averaging[42] and a linear regression line (solid gray), which approximates the data well over a wide range

of B values. The horizontal dashed line marks α_p^{exp}, and the vertical dashed line points at the value $B = 966\,\text{Å}^6$ kcal/mol, which is thus predicted to yield α_p^{exp}. This prediction then remains to be checked for the final model GP6P (cf. Table S4 below).

F. Additional Observables Computed for GP6P

In addition to the predictions calculated with GP6P for various important observables of liquid water at (p_0, T_0) and for the density profile $n(p_0, T)$ in the temperature range $T \in [250, 320]\,\mathrm{K}$ (see Sections III A 2 and III A 3), we here document the results obtained for the four target observables of the empirical parameter optimization, for extrapolations of size-dependent observables, for the water-water RDF, and for the microscopic electrostatic properties of the GP6P model.

1. Targeted Observables

Table S4 lists for the four target observables of the empirical optimization (cf. Sections S1 C 6 and S1 C 7) the experimental target values and the GP6P values obtained from 3 ns PMM-MD simulations by the methods described in Section S1 D. These observables are the average potential energy per molecule E_0, the position $r_{\mathrm{OO}}^{\mathrm{max}}$ of the first peak of the oxygen-oxygen RDF, the average pressure $\langle p \rangle$ and the thermal expansion coefficient α_p.

TABLE S4. Observables of the GP6P Model and the experimental values targeted by the parametrization. Errors were calculated by block averaging[42] from 3 ns PMM-MD trajectories.

quantity	unit	GP6P value	exp.	[Ref.]
E_0	kcal/mol	-9.920 ± 0.001	-9.92	[16]
$r_{\mathrm{OO}}^{\mathrm{max}}$	Å	2.757 ± 0.0005	2.76	[17]
$\langle p \rangle$	atm	1.8 ± 2.6	1	
α_p	10^{-4}/K	2.85 ± 0.1	2.8	[15]

Within the limits of statistical accuracy the calculated values exactly agree with their targets. This agreement indicates that the empirical optimization, which comprised the WC protocol described in Section S1 C 6 and the tuning of the dispersion parameter B (cf. Section S1 C 7) was actually successful.

2. Size-Dependent Observables

Three of the macroscopic observables, i.e. the self-diffusion constant D_0, the viscosity η, and the dielectric constant ϵ, relate to the infinite system and have to be extrapolated from $N_i V_i T_0$ simulations on the three finite-size systems N_i considered by us (see Ref. 1 for further explanations and references). These simulations and the methods for their evaluation are described in Section S1 D.

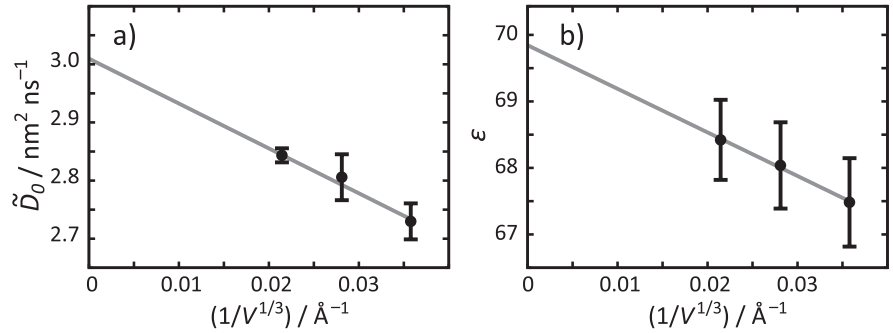


FIG. S15. Extrapolation graphs employed to determine (a) the diffusion constant D_0 and viscosity η and (b) the dielectric constant ϵ of the infinite system from PMM-MD simulations of finite size GP6P systems. The error bars represent standard errors of the mean.

Figures S15(a) and S15(b) show for the properly corrected self-diffusion constants $\tilde{D}_0(N_i, T_0)$ and for the iteratively calculated dielectric constants $\epsilon(N_i, T_0)$ linear extrapolation graphs as functions of the inverse system size $1/V^{1/3}$ (cf. Figures 8 and 9 in Ref. 1 and the associated discussion). At $1/V^{1/3} = 0$ they yield the values D_0 and ϵ of the infinite system, which are listed in Table II. The slope of the extrapolation line in Figure S15(a) is inversely proportional to the viscosity η (see the explanations in Ref. 1).

The error bars represent standard errors of the mean extracted from the sets of ten MD simulations underlying each data point. Error estimates for the extrapolated values and for η were obtained by dividing the 3×10 MD trajectories, which were employed to generate each of the two extrapolation graphs, into two halves, by executing for each of these halves an extrapolation, and by computing for the resulting small data sets the standard errors of the mean. These error estimates are also listed in Table II.

3. Water-Water Radial Distributions Function

Figure S16 compares the oxygen-oxygen RDF $g_{\mathrm{OO}}(r)$ computed for GP6P (black) with the experimental results of Soper[17] (gray). As explained in Section S1 C 6, the position

$r_{\mathrm{OO}}^{\mathrm{max,exp}} = 2.76\,\text{Å}$ of the first peak was one of the targets of the empirical optimization and, in this respect, the very close match of the GP6P RDF with the experimental finding was already documented by Table S4 above.

The height of the first peak, in contrast, and the overall shape of $g_{\mathrm{OO}}(r)$, are predictions of the GP6P model. This height is slightly underestimated and the remaining extrema are a little too pronounced. The positions of these extrema, however, agree quite well with the experimental data.

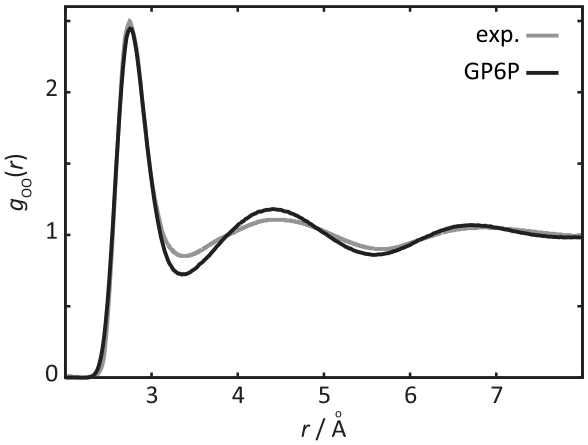


FIG. S16. Radial distribution function $g_{\mathrm{OO}}(r)$ in liquid water at (p_0, T_0) calculated for GP6P (black) and the experimental reference data of Soper (gray),[17] whose peak position at $r = 2.76\,\text{Å}$ was one of the targets of the parameterization.

4. Microscopic Electrostatic Properties of GP6P

The dipole and quadrupole moments of the water molecule are experimentally known only for the gas phase.[12,43] These data are given in the first line of Table S5 (because of $Q_{xx} + Q_{yy} + Q_{zz} = 0$, the quadrupole moment has only two non-trivial components). The second line characterizes the MT/BP description of the water molecule at the experimental gas-phase geometry[12,44] $G_{\mathrm{m}}^{\mathrm{g}} \equiv \{r_{\mathrm{OH}} = 0.9572\,\text{Å}, \varphi_{\mathrm{HOH}} = 104.52°\}$ and indicates that this DFT approach underestimates the polarity of the water molecule by about 3 %.

The DFT/PMM description of a solvated water molecule is provided by the fourth line of the table. Assuming that the static contributions to these moments suffer from the same 3 % underestimate as the MT/BP gas phase description one approximately arrives at the average electrostatic moments listed for GP6P in the last line of the table. Here we remark that also the induced dipole moment distribution centered around the oxygen atom provides a small contribution to the average GP6P quadrupole moments, because these moments are

TABLE S5. Components μ_x (in D) and Q_{yy}, Q_{zz} (in DÅ) of the total dipole and quadrupole moments, respectively, observed and calculated for an isolated water molecule in the gas phase geometry. Furthermore computational results are given for an isolated or solvated (averages) water molecule in the liquid phase geometry. The reference point is the center of mass.

	geometry	phase	μ_x	Q_{yy}	Q_{zz}
exp[12,43]	gas	gas	1.86	2.63	−2.50
MT/BP	gas	gas	1.80	2.56	−2.41
MT/BP	liq.	gas	1.79	2.64	−2.44
MT/BP	liq.	liq.	2.55	3.00	−2.83
GP6P	liq.	gas	1.86	3.02	−2.95
GP6P	liq.	liq.	2.54	3.07	−2.90

defined with respect to the center of mass of the water molecule. Thus, the GP6P quadrupole moments represent best estimates of the unknown liquid phase values.

The average liquid phase dipole moment $\langle \mu_{\mathrm{GP6P},x} \rangle = 2.544\,\mathrm{D}$ of GP6P appearing in the last line of Table S5 refers to the dipole distribution $p(\mu)$ depicted in Figure S17 as a green histogram and normal distribution, respectively. The standard deviation of this distribution is $\sigma(\mu^{\mathrm{i}}_{\mathrm{GP6P},x}) = 0.151\,\mathrm{D}$. For comparison the figure additionally shows the normal distribution (red) of the x-components of the total dipole moments $\boldsymbol{\mu}_{\mathrm{DFT}}$, which were calculated for the DFT fragments in the final ensemble $\mathcal{S}^6$ of the parameterization cycle. The associated distribution of induced DFT dipole moments was shown already in Figure S12 above. The two distributions are almost identical, because GP6P is constructed (cf. Section S1 C 5)

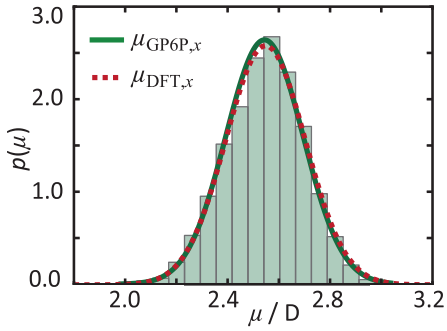


FIG. S17. The normalized histogram $p(\mu)$ and the associated normal distribution resulting for the components $\mu_{\mathrm{GP6P},x}$ of the GP6P dipole moments in the liquid phase (green) are compared with the normal distribution (red) of the component $\mu_{\mathrm{DFT},x}$ of the DFT dipole moments in the snapshot ensemble $\mathcal{S}^6$.

from a DFT-adapted model by replacing the DFT values ($\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{DFT}}, \alpha^{\mathrm{g}}_{\mathrm{DFT}}$) by their experimental counterparts ($\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{exp}}, \alpha^{\mathrm{g}}_{\mathrm{exp}}$), which enhances the static polarity by 3.5 % and reduces the polarizability by 8.3 %. Because the induced dipole moment amounts to only 30 % of the total dipole moment the two effects roughly compensate each other.

S2. ADDITIONAL RDFS FOR NMA IN GP6P SOLUTION

Here, we provide evidence that the cost-effective mean field (MFMD) approach yields the same RDFs of GP6P water in the surroundings of NMA as straight DFT/PMM-MD simulations. Subsequently we show that the optimal LJ parameter set $\mathcal{L}^{\mathrm{opt}}$, which has been determined for the core atoms of NMA's AG with the MT/BLYP setup, is equally well applicable to the MT/BP and MT/B3LYP setups.

A. RDFs from MD and MFMD are Equivalent

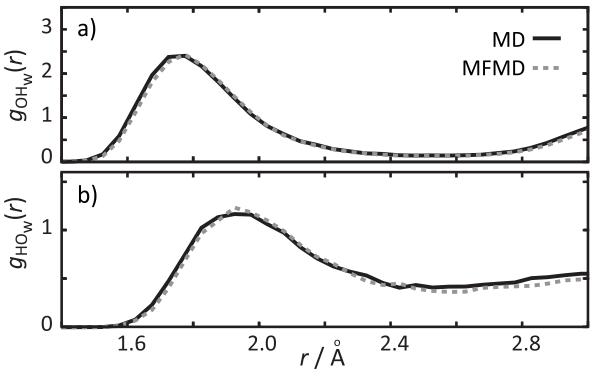


FIG. S18. RDFs $g_{\mathrm{OH_w}}(r)$ (a) and $g_{\mathrm{HO}}(r)$ (b) obtained from DFT/PMM-MFMD (dotted), and DFT/PMM-MD (solid).

Figure S18 shows RDFs a) $g_{\mathrm{OH_w}}(r)$ and b) $g_{\mathrm{HO_w}}(r)$ of GP6P water around the C′=O and N–H groups, respectively, of the DFT fragment NMA. Here, solid and dotted lines depict results of differently expensive DFT/PMM simulations carried out with the MT/BLYP setup and the LJ parameter set $\mathcal{L}^{\mathrm{opt}}$. The solid lines mark the RDFs from usual DFT/PMM-MD simulations (cf. Section II E 3), whereas the dotted curves represent the results of the inexpensive MFMD approach introduced in Section II F 2.

For both hydrogen bonding sites the agreement between the RDFs from usual MD simulations and from the cost-effective MFMD approach is very good. Therefore MFMD is well-suited for the computationally demanding iterative optimization of LJ parameters for DFT atoms in a hybrid simulation system.

B. RDFs for Different DFT Setups

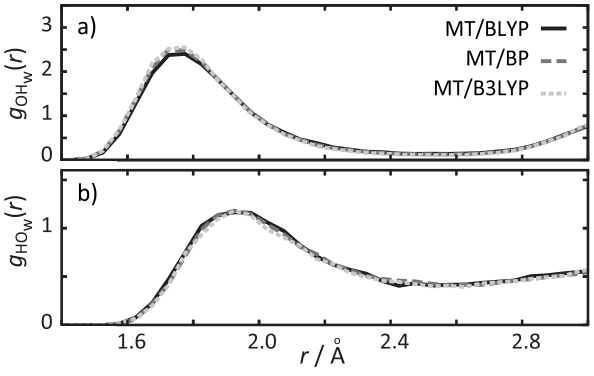


FIG. S19. RDFs $g_{\mathrm{OH_w}}(r)$ (a) and $g_{\mathrm{HO}}(r)$ (b) obtained from DFT/PMM-MD employing the MT/BLYP (solid), MT/BP (dashed), and MT/B3LYP (dotted) setups for the parameter set $\mathcal{L}^{\mathrm{opt}}$.

We have claimed in Section III B that RDFs of GP6P water around the AG of the DFT fragment NMA do hardly depend on the chosen DFT setup. Figure S19 proves for the three alternative setups MT/BLYP (solid), MT/BP (dashed), and MT/B3LYP (dotted) that this is actually the case. The deviations of the associated RDFs depicted in a) and b), which belong to the two NMA–water hydrogen bonding sites, are very small. Thus, the LJ parameter sets of the DFT fragments, which can strongly influence such RDFs, are quite well transferable between different DFT setups.

S3. ADDITIONAL IR SPECTRA OF NMA

In addition to the gas and aqueous phase IR spectra in the spectral range [1800, 1000] cm^{-1}, which are displayed by Figure 6 for the MT/B3LYP setup, we here present the corresponding spectra also for the MT/BLYP and MT/BP setups. As far as the positions of the six main peaks in the aqueous phase spectra are concerned, the added data solely repeat the contents of Figure 7. Furthermore, we show and discuss the low-frequency IR spectra covering the range [1380, 580] cm^{-1} for all three DFT setups. Note that Table S6 at the bottom of this section gathers all calculated AI-AVI frequencies.

A. IR Spectra of NMA between 1800 cm^{-1} and 1000 cm^{-1}

Figures S20(a) and S20(b) compare the DFT-MD and DFT/PMM-MD predictions of NMA's gas and aqueous phase IR spectra, respectively, derived with the three different DFT setups MT/BLYP, MT/BP, and MT/B3LYP. All frequencies have been scaled by the

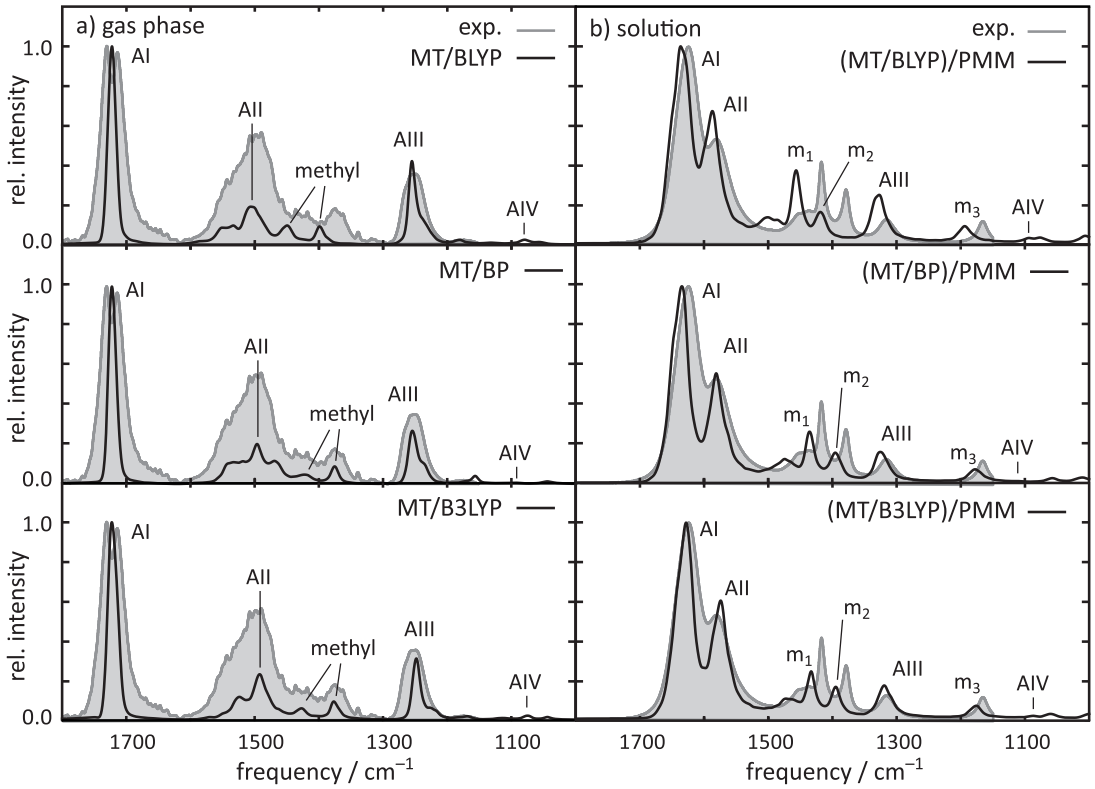


FIG. S20. The IR spectrum of NMA in a) the gas phase and b) aqueous solution as determined by IR spectroscopy[45] (gray) and by FTTCF (black) from DFT-MD and DFT/PMM-MD trajectories employing the MT/BLYP, MT/BP, and MT/B3LYP setups. Frequencies were scaled with the respective gas phase scaling factors ($f_{\mathrm{MT/BLYP}} = 1.0428$, $f_{\mathrm{MT/BP}} = 1.0335$, $f_{\mathrm{MT/B3LYP}} = 0.9988$); the heights of the AI peaks were normalized to one. Band assignments were obtained by a GNC analysis.[46,47]

setup-specific gas phase scaling factors f_{DFT}. The corresponding experimental data[45] are shown in gray. The MT/B3LYP results at the bottom were already presented in Sections III C 3 and III C 4 of the main text and are repeated here for visual comparison.

1. Gas Phase: MT/BLYP vs. MT/BP

We have claimed in Section III C 3 that the scaled MT/BLYP and MT/BP setups provide a likewise excellent description of the three amide bands as the MT/B3LYP setup and that solely MT/BLYP shows a substantially poorer performance on the methyl bands. An inspection of Figure S20(a) immediately verifies these claims.

While marginally scaled B3LYP predicts the AII and AIII bands at 1492 cm^{-1} and

1248 cm^{-1}, respectively, scaled MT/BLYP yields 1505 cm^{-1} and 1255 cm^{-1}, and scaled MT/BP 1497 cm^{-1} and 1255 cm^{-1}. For each of the two amide bands these frequencies are close to each other and to the experimental data. Looking at the methyl bands one immediately recognizes the large blue-shifts in the BLYP prediction at the top of Figure S20(a). Here, the predicted peaks at 1450 cm^{-1} and 1399 cm^{-1} overestimate the spectroscopic data by about 25 cm^{-1}. MT/BP-MD, in contrast, predicts bands at 1423 cm^{-1} and 1376 cm^{-1}, which, just like the MT/B3LYP results, closely agree with the experimental findings.

2. Aqueous Solution: MT/BLYP vs. MT/BP

Figure S20(b) provides a visual support for the comparison of peak frequencies presented by Figure 7. This comparison has shown that the scaled MT/BP and MT/BLYP setups describe the amide band frequencies of NMA in water almost as well as the marginally scaled MT/B3LYP setup. For the methyl bands, however, the MT/BLYP setup predicted much larger frequency overestimates than the other two setups. Thus, particularly MT/BP outperforms MT/BLYP also for the aqueous phase.

But despite these details, all calculated spectra reproduce the observed spectral features quite well. In combination with the GNC analysis the calculations therefore enable a clear-cut assignment of the observed bands to local modes.[46,47]

B. Low-Frequency IR Spectra

In addition to the IR spectrum of NMA in the range $[1800, 1000]$ cm^{-1}, which we have discussed for the gas and aqueous phases (cf. Sections III C 3, III C 4, and S3 A), here we show the corresponding data for the low-frequency range down to 580 cm^{-1}. Figure S21 presents for the three considered DFT setups the GNC[46,47] decompositions of the IR spectra (black) calculated for NMA in the gas (a) and aqueous (b) phases by DFT-MD and DFT/PMM-MD, respectively. The frequencies were scaled with the setup-specific gas phase scaling factors f_{DFT}. Four amide bands were identified by the GNC analysis in the depicted spectral range, namely AIII (red), AIV (green), AV (blue), and AVI (orange). Graphical representations of these AG normal modes are provided by Figure 7 in Ref. 48. The frequencies calculated for these bands are listed in Table S6.

All three DFT setups predict for the low-frequency amide bands solvatochromic blue-shifts upon transfer of NMA from the gas into the aqueous phase. For the AIV, AV and AVI bands these blue-shifts amount on average to 14 cm^{-1}, 45 cm^{-1}, and 21 cm^{-1}, respectively. Among the DFT setups the calculated frequencies slightly differ with scaled MT/BP

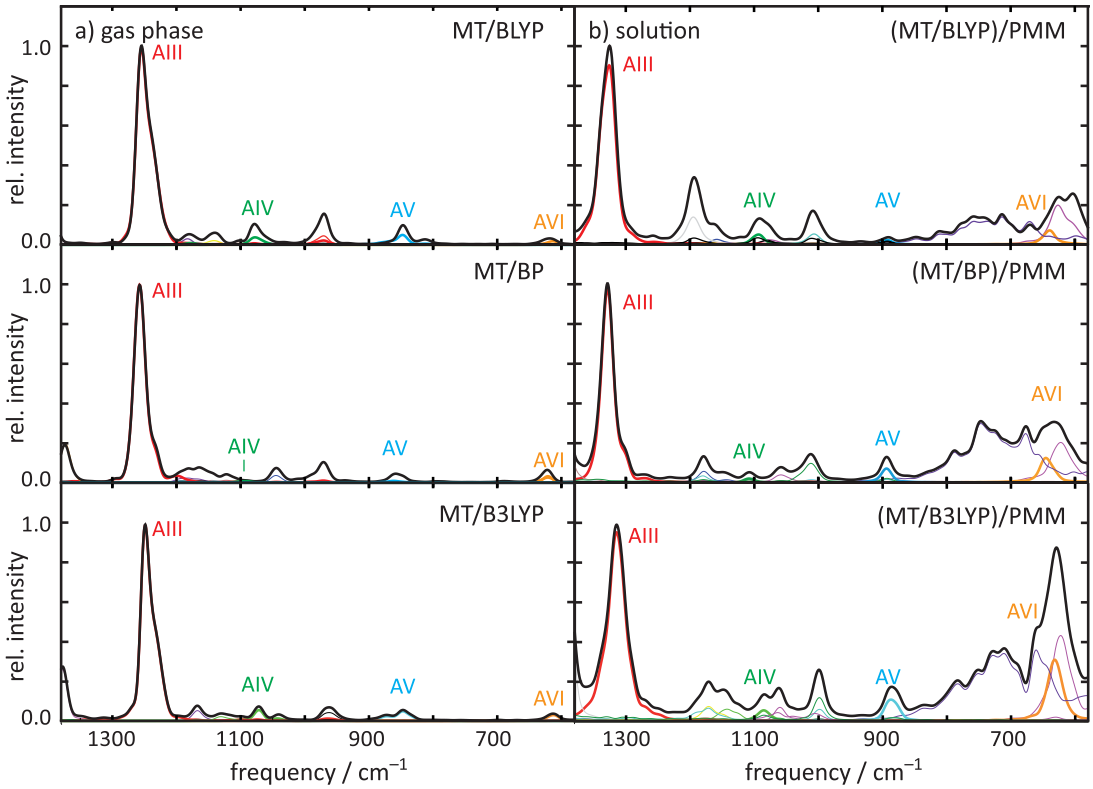


FIG. S21. Decomposition of the IR spectra (black) of NMA in the frequency region below 1380 cm^{-1} by the GNC[46,47] analysis for a) gas phase DFT-MD and b) condensed phase DFT/PMM-MD using the MT/BLYP, MT/BP, or MT/B3LYP setups. Local contributions by the AIII, AIV, AV, and AVI modes are drawn as thick lines in red, green, blue, and orange, respectively. Frequencies were scaled with the respective gas phase scaling factors f_{DFT}; the heights of the AIII peaks were normalized to one.

generally predicting the highest and scaled MT/B3LYP the lowest frequency values.

TABLE S6. Amide band frequencies of NMA in vacuum and aqueous solution (units: cm^{-1}). All DFT values were scaled with the respective setup-specifc gas phase scaling factor ($f_{MT/BLYP} = 1.0428$, $f_{MT/BP} = 1.0335$, $f_{MT/B3LYP} = 0.9988$).

	exp.[45]		MT/BLYP		MT/BP		MT/B3LYP	
	vac.	sol.	vac.	sol.	vac.	sol.	vac.	sol.
AI	1723	1625	1723	1634	1723	1635	1723	1628
AII	1499	1582	1505	1587	1497	1581	1492	1574
AIII	1255	1317	1255	1328	1255	1326	1248	1319
AIV	–	–	1079	1093	1093	1107	1071	1085
AV	–	–	842	892	849	893	847	888
AVI	–	–	618	640	623	645	612	632

REFERENCES

[1]P. Tröster, K. Lorenzen, M. Schwörer, and P. Tavan, J. Phys. Chem. B **117**, 9486 (2013).
[2]P. Tröster, K. Lorenzen, and P. Tavan, J. Phys. Chem. B **118**, 1589 (2014).
[3]P. Tröster and P. Tavan, J. Phys. Chem. Lett. **5**, 138 (2014).
[4]M.-L. Tan, J. R. Cendagorta, and T. Ichiye, J. Chem. Phys. **141**, 244504 (2014).
[5]M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan, and G. Mathias, J. Chem. Phys. **138**, 244103 (2013).
[6]M. Schwörer, K. Lorenzen, G. Mathias, and P. Tavan, J. Chem. Phys. **142**, 104108 (2015).
[7]K. Lorenzen, C. Wichmann, and P. Tavan, J. Chem. Theory Comput. **10**, 3244 (2014).
[8]K. Lorenzen, G. Mathias, and P. Tavan, J. Chem. Phys. **143**, 184114 (2015).
[9]K. Ichikawa, Y. Kameda, T. Yamaguchi, H. Wakita, and M. Misawa, Mol. Phys. **73**, 79 (1991).
[10]W. E. Thiessen and A. H. Narten, J. Chem. Phys. **77**, 2656 (1982).
[11]B. Schropp and P. Tavan, J. Phys. Chem. B **114**, 2051 (2010).
[12]S. Clough, Y. Beers, G. Klein, and L. Rothman, J. Chem. Phys. **59**, 2254 (1973).
[13]W. F. Murphy, J. Chem. Phys. **67**, 5877 (1977).
[14]R. A. Buckingham and J. Corner, Proc. R. Soc. London, Ser. A **189**, 118 (1947).
[15]G. S. Kell, J. Chem. Eng. Data **12**, 66 (1967).
[16]G. Jancso and W. A. Van Hook, Chem. Rev. **74**, 689 (1974).
[17]A. K. Soper, Chem. Phys. **258**, 121 (2000).
[18]IPHIGENIE is available for download free of charge under the GPL licence at `http://sourceforge.net/projects/iphigenie`.
[19]J. Hutter, A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman, and M. Parrinello, *CPMD: Car–Parinello Molecular Dynamics, version 3.17.1*, © IBM Corp 1990–2008 and MPI für Festkörperforschung Stuttgart 1997–2001, `www.cpmd.org`.
[20]G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126** (2007).
[21]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[22]M. P. Allen and D. Tildesley, *Computer Simulations of Liquids* (Clarendon, Oxford, 1987) Chap. 2.4, pp. 46–50.
[23]D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge university press, Cambridge, UK, 2004) Chap. 2.3, pp. 18–20.
[24]B. Schropp and P. Tavan, J. Phys. Chem. B **112**, 6233 (2008).
[25]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[26]J. Perdew and W. Yue, Phys. Rev. B **33**, 8800 (1986).
[27]N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[28]Refs. 1 and 2 contain a misprint concerning $\alpha^{\mathrm{g}}_{\mathrm{DFT}}$. Instead of the correct MT/BP value[11] 1.6025 Å^3 belonging to the experimental liquid phase geometry G_{m} one reads 1.58 Å^3, which is[11,24] the MT/BP polarizability at the MT/BP vacuum geometry.

[29]K. Levenberg, J. Numer. Anal. **16**, 588 (1944).

[30]D. W. Marquardt, Journal of the Society for Industrial & Applied Mathematics **11**, 431 (1963).

[31]C. Berweger, W. van Gunsteren, and F. Müller-Plathe, Chem. Phys. Lett. **232**, 429 (1995).

[32]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[33]M. Lingenheil, R. Denschlag, R. Reichold, and P. Tavan, J. Chem. Theory Comput. **4**, 1293 (2008).

[34]K. Krynicki, C. Green, and D. Sawyer, Faraday Discuss. Chem. Soc. **66**, 199 (1978).

[35]P. Tröster, K. Lorenzen, M. Schwörer, and P. Tavan, J. Phys. Chem. B **118**, 13321 (2014).

[36]P. E. Smith and W. F. van Gunsteren, J. Chem. Phys. **100**, 3169 (1994).

[37]G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).

[38]R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).

[39]Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[40]T. Okabe, M. Kawata, Y. Okamoto, and M. Mikami, Chem. Phys. Lett. **335**, 435 (2001).

[41]M. Lingenheil, R. Denschlag, G. Mathias, and P. Tavan, Chem. Phys. Lett. **478**, 80 (2009).

[42]D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge university press, Cambridge, UK, 2004) Chap. 4.2, pp. 84–89.

[43]J. Verhoeven and A. Dymanus, J. Chem. Phys. **52**, 3222 (1970).

[44]W. S. Benedict, N. Gailar, and E. K. Plyler, J. Chem. Phys. **24**, 1139 (1956).

[45]J. Kubelka and T. A. Keiderling, J. Phys. Chem. A **105**, 10922 (2001).

[46]G. Mathias and M. Baer, J. Chem. Theory Comput. **7**, 2028 (2011).

[47]G. Mathias, S. D. Ivanov, A. Witt, M. D. Baer, and D. Marx, J. Chem. Theory Comput. **8**, 224 (2011).

[48]V. Schultheis, R. Reichold, B. Schropp, and P. Tavan, J. Phys. Chem. B **112**, 12217 (2008).

2.4 Effizientes Abtasten von Konformationsräumen mit DFT/PMM-MD

Für die Untersuchung langsamer Prozesse sind konventionelle MD-Simulationen aufgrund der relativ kurzen maximal erreichbaren Zeitskalen nur von limitiertem Nutzen. Zur Erhöhung der Effizienz sind daher sog. generalisierte Ensemblemethoden [196, 197] entwickelt worden, die z.B. das Abtasten der in Abschnitt 1.1.3 beschriebenen Konformationsräume von Polypeptiden erheblich beschleunigen können.

Ein generalisiertes Ensemble wird dabei von K Kopien k ($k = 0, 1, \ldots, K-1$) eines Simulationssystems aufgespannt, die sich z.B. durch die mittlere Temperatur T_k oder eine modifizierte Energiefunktion U_k systematisch so voneinander unterscheiden, dass die Abtastrate der jeweiligen relevanten Koordinaten erhöht wird. So können z.B. durch Temperaturerhöhung freie Energiebarrieren effektiv erniedrigt und die Rate an Konformationsübergängen erhöht werden, was die Grundlage von Algorithmen wie *simulated tempering* (ST) [198, 199] oder *replica exchange* (RE) [200, 201] bildet.

Da die Effizienz von Verfahren wie ST oder RE von der Zahl der behandelten Freiheitsgrade abhängt, wurden Ansätze wie *solute tempering* [202, 203] vorgeschlagen, die die Abtastrate effektiv nur für einen Ausschnitt des Simulationssystems (z.B. ein Polypeptid) erhöhen, und die Zahl der behandelten Freiheitsgrade damit reduzieren. Denschlag *et al.* entwickelten durch geschicktes Zusammenführen von simulated tempering und solute tempering die generalisierte Ensemblemethode SST [195], die einen weiteren Effizienzgewinn verspricht.

Die nachfolgende abgedruckte Publikation

Simulated Solute Tempering in Fully Polarizable
Hybrid QM/MM Molecular Dynamics Simulations

Magnus Schwörer, Christoph Wichmann, Erik Gawehn und Gerald Mathias
J. Chem. Theory Comput. **12**, 992-999 (2016),

die ich zusammen mit Christoph Wichmann, Erik Gawehn, und Gerald Mathias verfasst habe, zeigt, dass das für unpolarisierbare MM-Kraftfelder entwickelte SST-Verfahren [195] auf PMM- und DFT/PMM-Systeme anwendbar ist. Hierbei können die für DFT/PMM-SST benötigten Gewichtsparameter w_k mittels einer einfachen Rechenvorschrift aus effizienten PMM-SST-Rechnungen vorhergesagt werden. Als Anwendungsbeispiel wird die freie Energielandschaft eines DFT-Modells von Alanindipeptid in PMM-Wasser berechnet.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

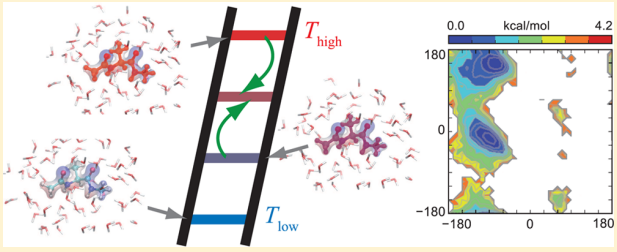
Simulated Solute Tempering in Fully Polarizable Hybrid QM/MM Molecular Dynamics Simulations

Magnus Schwörer, Christoph Wichmann, Erik Gawehn, and Gerald Mathias*

Lehrstuhl für BioMolekulare Optik, Ludwig−Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

Supporting Information

ABSTRACT: We successfully apply a solute tempering approach, which substantially reduces the large number of temperature rungs required in conventional tempering methods by solvent charge scaling, to hybrid molecular dynamics simulations combining quantum mechanics with molecular mechanics (QM/MM). Specifically, we integrate a combination of density functional theory (DFT) and polarizable MM (PMM) force fields into the simulated solute tempering (SST) concept. We show that the required DFT/PMM-SST weight parameters can be obtained from inexpensive calculations and that for alanine dipeptide (DFT) in PMM water three rungs suffice to cover the temperature range from 300 to 550 K.



1. INTRODUCTION

The sampling of the conformational space of (bio)molecules by plain molecular dynamics (MD) simulations is computationally inefficient. If one aims at an unbiased structural ensemble or has no prior knowledge of the free energy landscape, so-called generalized ensemble[1] methods can alleviate the sampling problem. Many variants have been proposed over the past years, among which simulated tempering (ST)[2,3] and temperature replica exchange (RE)[4,5] are most popular examples.

In ST, a single copy (replica) of the system is simulated at a temperature T_i, which belongs to the rung $i \in \{0,1,..., R-1\}$ of a ladder covering the temperatures $\mathcal{T} = \{T_0, T_1, ..., T_{R-1}\}$ in increasing order. Periodically, the replica tries to switch from the rung i to a new rung $j \neq i$ at $T_j \in \mathcal{T}$ with a probability $p_{ij} = \min[1,\exp(\Delta_{ij})]$, which is given by a Metropolis criterion[6] serving to preserve the statistical ensemble at each rung. Typically, the target rung $j = i \pm 1$ is chosen by a stochastic or deterministic scheme.[7] The general ST form[2,3] of the exponent

$$\Delta_{ij} = [\beta_i E_i(\mathbf{R}) - w_i] - [\beta_j E_j(\mathbf{R}) - w_j] \quad (1)$$

in the Metropolis criterion depends on the inverse temperatures $\beta_k = (k_B T_k)^{-1}$ (k_B is the Boltzmann constant), on rung-dependent potential energies $E_k(\mathbf{R})$ of the current system configuration $\mathbf{R} = (\mathbf{r}_1, ..., \mathbf{r}_N)^T$ of the N atoms at the positions $\mathbf{r}_n$ ($1 \leq n \leq N$) of the simulation system, and on dimensionless weights w_k, which may be freely chosen, e.g., to sample all rungs uniformly. Upon changing a replica's rung from i to j, the momenta of all N atoms are scaled by $\sqrt{T_j/T_i}$ to correspond to the new target temperature T_j.[5] Note that in conventional ST, $E_k(\mathbf{R}) \equiv E(\mathbf{R})$ is independent of the rung k.

The number R of required temperature rungs grows with the number of degrees of freedom (DOF) of the tempered system. Larger ladders require longer simulations or more replicas that are run in parallel[8] to acquire the same amount of sampling statistics at a given rung. In this respect ST is advantageous because it generally needs about 30% less rungs than RE[9] but at the expense of requiring the *a priori* unknown weight parameters w_k. In most cases, however, the w_k can be determined with a reasonably limited effort,[8,10] and several schemes have been proposed for the *a priori* determination and/or on-the-fly update of optimal weights.[8,10−15]

Particularly, if one simulates multiple independent ST replicas in parallel, their data can be combined to quickly converge the w_k.[8,15]

If one is interested only in the ensemble at a particular temperature, e.g., at ambient conditions, the solute tempering approach[16] can drastically decrease the ladder size for both RE (RE with solute tempering, REST)[16] and ST (simulated solute tempering, SST)[10] generalized ensemble simulations. Solute tempering splits the potential energy

$$E_k(\mathbf{R}) = \lambda_k^{pp} E^{pp}(\mathbf{R}) + \lambda_k^{ps} E^{ps}(\mathbf{R}) + \lambda_k^{ss} E^{ss}(\mathbf{R}) \quad (2)$$

of the system at rung k into a linear combination of the solute−solute interaction energy E^{pp}, solute−solvent energy E^{ps}, and solvent−solvent energy E^{ss}. Choosing the linear coefficients[10,16]

$$\lambda_k^{pp} = 1, \quad \lambda_k^{ps} = \gamma_k, \text{ and } \lambda_k^{ss} = \gamma_k^2 \quad (3)$$

with $\gamma_k = \sqrt{\beta_0/\beta_k}$ and inserting eqs 2 and 3 in eq 1 yields the SST exponent[10]

Received: October 6, 2015
Published: February 2, 2016

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

$$\Delta_{ij} = (\beta_i - \beta_j)E^{pp}(\mathbf{R}) + (\sqrt{\beta_0\beta_i} - \sqrt{\beta_0\beta_j})E^{ps}(\mathbf{R}) - (w_i - w_j) \quad (4)$$

Because of the chosen prefactor $\lambda_k^{ss} = \gamma_k^2 = \beta_0/\beta_k$ for the solvent−solvent energy E^{ss} in eq 2, its contribution cancels in the derivation of eq 4. Consequently, the solvent's DOF no longer enter the exchange probability p_{ij}, and the number of temperature rungs can be largely reduced, which is the key idea of solute tempering.[10,16] Note that for REST and SST, one can formulate an equivalent Hamiltonian RE[17,18] or ST scheme by using the modified factors $\tilde{\lambda}_k^x = \gamma_k^{-2}\lambda_k^x$, $x \in \{pp, ps, ss\}$ for E_k (2). In such a so-called solute scaling scheme,[19] one simulates at the reference temperature T_0 on all rungs k and does not have to scale momenta or to modify thermostats upon rung changes.[19−21]

For $k = 0$, all linear coefficients (eq 3) become unity. At this lowest rung, eq 2 thus represents the unmodified potential energy $E_0(\mathbf{R}) \equiv E(\mathbf{R})$, and the canonical ensemble at T_0 is sampled. At higher rungs, $k > 0$, $E_k(\mathbf{R})$ is modified, and the sampled ensembles have no relevant physical interpretation. For plain molecular mechanics (MM) force fields, which exclusively employ pairwise additive long-range interactions, the linear coefficients (eq 3) needed for E_k simply follow from scaling all solvent charges by γ_k as suggested in ref 10. This scaling considerably simplifies the implementation of SST because the long-range electrostatic contributions can be calculated all at once by Ewald or fast multipole type routines and do not have to be split into solute−solute, solute−solvent, and solvent−solvent contributions in order to scale these separately. For van der Waals interactions, one can simply scale the parameters of the pair potentials.[10]

A high sampling efficiency is particularly important if one employs hybrid methods combining quantum mechanical (QM) methods and MM force fields[22−24] because QM/MM-MD simulations require a much larger computational effort than plain MM-MD. This effort is particularly large if a higher-level method such as density functional theory[25,26] (DFT) is applied to the QM subsystem.[27−30] Correspondingly, only a few combinations of QM/MM and generalized ensemble techniques like RE have been reported,[31−34] which had to cope with the large number of rungs required for condensed phase systems.

Using solute tempering approaches like REST[16] or SST[10] for QM/MM simulations to reduce the number of rungs is an obvious and favorable alternative because the required partitioning into solute and solvent subsystems is naturally given by selecting the QM and MM subsystems as the solute and solvent, respectively. Unfortunately, however, combining solute tempering with QM/MM is not straightforward because in contrast to a MM setup the QM/MM Hamiltonian is not pairwise additive but contains many-body interactions. They prevent the potential energy function E from being easily separated into the contributions E^{pp}, E^{ps}, and E^{ss} required to write E_k in the form of eq 2. The issue of the lacking separability is exacerbated if instead of a plain MM solvent model a polarizable MM (PMM) solvent force field is employed, since here many-body contributions are additionally contained in the solvent−solvent interactions.

In this article, we will demonstrate that solute tempering with charge scaling can nevertheless be readily extended to systems with many-body interactions. Here, we will specifically consider a recently presented DFT/PMM hybrid method, which has been shown to describe solute−solvent interactions at a high accuracy.[30,35] The method enables a rapid and accurate computation of the DFT/PMM electrostatic interaction Hamiltonian

$$H_{DFT/PMM} = \int dr^3 \rho(\mathbf{r})\Phi(\mathbf{r}; \mathbf{Q},\mathbf{P}) \quad (5)$$

which represents the energy of the DFT solute's charge density $\rho(\mathbf{r})$ in the external PMM potential $\Phi(\mathbf{r}; \mathbf{Q},\mathbf{P})$. The sources of Φ are static partial charges $\mathbf{Q} = (q_1, ..., q_M)^T$ and induced Gaussian dipoles $\mathbf{P} = (\mathbf{p}_1, ..., \mathbf{p}_M)^T$, which reside at the M solvent atoms $s = 1,...,M$ and polarize $\rho(\mathbf{r}) \equiv \rho(\mathbf{r}; \mathbf{Q},\mathbf{P})$. The strengths

$$\mathbf{p}_s = \alpha_s \mathbf{E}(\mathbf{r}_s; \mathbf{Q},\mathbf{P}, \rho) \quad (6)$$

of the induced atomic dipole distributions depend via a scalar polarizability α_s linearly on the polarizing electric field $\mathbf{E}(\mathbf{r}_s)$ generated at their centers $\mathbf{r}_s$ by all other electrostatic sources in the system. Therefore, $\mathbf{P}$ and ρ have to be iteratively determined until a joint self-consistency is reached in each MD integration step.[30]

2. THEORY

For the use of solute tempering in polarizable simulation settings like DFT/PMM, it is a seemingly reasonable idea to apply the same scaling concept as in plain MM force field simulations. This is obvious for the van der Waals interactions between QM and MM fragments because they are usually treated by the additive pair-expressions of a MM force field, which is why we will omit these contributions in the subsequent discussion. The effect of the solvent charge scaling $\mathbf{Q} \rightarrow \gamma_k\mathbf{Q}$ in DFT/PMM-SST simulations is, however, not likewise obvious.

Because of the mutual polarization between the solute and the solvent, their interaction energy $E_k^{ps} = E_k - E_k^{pp} - E_k^{ss}$ at a rung k has to be computed from the difference of the potential energy E_k of the whole system and the solute−solute and solvent−solvent interaction energies E_k^{pp} and E_k^{ss}, respectively. If polarizable DOF are present, these latter two energies must be calculated for the *isolated* solute subsystem and the *isolated* solvent subsystem, respectively, to exclude the polarization between the fragments. E_k^{ps} then collects all those contributions to E_k, which arise from solute−solvent interactions (eq 5), and the additional solute−solute and solvent−solvent interaction energies, which are due to mutual solute−solvent polarization effects.

For the isolated PMM solvent subsystem containing partial charges $\mathbf{Q}$ and induced dipoles $\mathbf{P}$, it is straightforward to verify that for scaled charges $\mathbf{Q} \rightarrow \gamma_k\mathbf{Q}$ the linear response eq 6 yields the scaled dipoles $\gamma_k\mathbf{P}$ as the self-consistent solution. Because all associated electrostatic energy contributions are proportional[30] to products $q_sq_{s'}$, $q_s\mathbf{p}_{s'}$, or $\mathbf{p}_s\mathbf{p}_{s'}$, the solvent charge scaling yields the desired prefactor $\lambda_k^{ss} = \gamma_k^2$ for the potential energy $E_k^{ss} = \gamma_k^2E_0^{ss} \equiv \gamma_k^2E^{ss}$ of the isolated solvent subsystem (cf. eq 3). Trivially, the potential energy of the isolated solute subsystem is independent of γ_k, and the prefactor λ_k^{pp} is unity, i.e., $E_k^{pp} = E_0^{pp} \equiv E^{pp}$.

The remaining SST prefactor λ_k^{ps} of $E_0^{ps} \equiv E^{ps}$ is undetermined for polarizable setups like DFT/PMM because it depends on the mutual polarization between the two subsystems, which is nonlinear in γ_k. In analogy to eq 2, we therefore express the potential energy of the whole system as

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

$$E_k(\mathbf{R}) = E^{pp}(\mathbf{R}) + [\gamma_k + \epsilon_k(\mathbf{R})]E^{ps}(\mathbf{R}) + \gamma_k^2 E^{ss}(\mathbf{R}) \quad (7)$$

Here, the prefactor λ_k^{ps} of E^{ps} is split into the MM-SST prefactor γ_k and a configuration-dependent quantity

$$\epsilon_k(\mathbf{R}) = \frac{E_k^{ps}(\mathbf{R}) - \gamma_k E^{ps}(\mathbf{R})}{E^{ps}(\mathbf{R})} \quad (8)$$

which is nonzero for $k > 0$.

The additional term $\epsilon_k(\mathbf{R})$ covers two effects. First, the solute contributions to the polarizing field entering eq 6 is in leading order independent of the solvent scale factors γ_k, whereas the polarizing field generated by the solvent charges scales linearly with γ_k. Therefore, induced solvent dipoles, which are located in the vicinity of the solute, will scale with γ_k^α, where $\alpha < 1$. Second, solvent charge scaling also changes the polarization of the solute. This extra polarization also contributes nonlinearly in γ_k to the solute−solvent interaction energy. Note that $\epsilon_k(\mathbf{R})$ vanishes if neither the solute nor the solvent are polarizable. If the added terms $\epsilon_k(\mathbf{R})$ are small compared to the respective γ_k, their presence in the prefactor of E^{ps} (cf. eq 7) should have only a minor effect on the efficiency of DFT/PMM-SST.

Inserting eq 7 into eq 1 yields the exponent

$$\Delta_{ij} = (\beta_i - \beta_j)E^{pp}(\mathbf{R}) + \left[\sqrt{\beta_0\beta_i} - \sqrt{\beta_0\beta_j} + \beta_i\epsilon_i(\mathbf{R}) - \beta_j\epsilon_j(\mathbf{R})\right]E^{ps}(\mathbf{R}) - (w_i - w_j) \quad (9)$$

which, like eq 4, does not depend on E^{ss}. Thus, the key idea of solute tempering,[16] which is the removal of the solvent−solvent energy E^{ss} from Δ_{ij}, is preserved by solvent charge scaling[10] also in polarizable simulation systems. Like in eq 7, also in eq 9, the prefactor of E^{ps} is modified by configuration-dependent contributions $\beta_k\epsilon_k(\mathbf{R})$.

Note that one does not have to evaluate the Metropolis exponent Δ_{ij} through eq 9 if one wants to execute an exchange step in a SST simulation. Instead one computes the potential energies $E_i(\mathbf{R})$ and $E_j(\mathbf{R})$ of the current configuration $\mathbf{R}$ simply by scaling the solvent charges with γ_i and γ_j, respectively. With these energies, Δ_{ij} is easily obtained from the much simpler eq 1. This particularly implies that no calculations on the isolated solute and/or solvent subsystems are necessary during a DFT/PMM-SST simulation in order to obtain E^{pp}, E^{ps}, and E^{ss} separately. Further below, we will conduct such separate calculations only to investigate the quantity ϵ_k in detail.

Because E^{pp} and E^{ps} are not readily separated in DFT/PMM settings, the solute scaling approach[19−21] mentioned above would also have to resort to the charge scaling method. Using on rung k scaled charges $\gamma_k\mathbf{Q}$, and concurrently scaling down the resulting *total* forces and energies of *all* atoms by γ_k^{-2} yields the desired scaling factors $\tilde{\lambda}_k^{pp} = \gamma_k^{-2}$, $\tilde{\lambda}_k^{ps} \approx \gamma_k^{-1}$, and $\tilde{\lambda}_k^{ss} = 1$, and thus renders an implementation of solute scaling also for DFT/PMM Hamiltonians.

The above considerations on DFT solutes equally apply to any other *polarizable solute* model, e.g., QM or PMM, and any other implementation of *solvent polarizability*, which depends linearly on the polarizing field (cf. eq 6). Furthermore, the scaling approach[10] can be equally used to combine REST with QM/(P)MM hybrid simulations.

3. METHODS

The sample systems, by means of which we examine the properties of DFT/PMM-SST, are DFT and MM models of alanine dipeptide (Ac-Ala-NHMe) in PMM water. For the latter, we chose a six-point water model featuring five Gaussian static charges and a single Gaussian induced dipole. This PMM water potential has been specifically developed for accurate DFT/PMM-MD simulations[36] of molecules in aqueous solution. The details of the simulations (including the parameters of the water model) are thoroughly described in Sections S1−S3 of the Supporting Information (SI).

All simulations were performed with the IPHIGENIE/CPMD program package, which combines the parallel PMM-MD code IPHIGENIE[30,35,37−41] with the parallel grid-based plane-wave DFT code CPMD.[42] Reasonable initial conditions for the DFT/PMM simulations were prepared by executing much less costly MM/PMM-SST-MD simulations, in which the alanine dipeptide solute was described by its CHARMM22/CMAP[43,44] model. Here, a first SST ensemble $\mathcal{E}$ was generated by distributing $C = 32$ replicas on the four-rung temperature ladder $\mathcal{T} = \{300\text{ K}, 367\text{ K}, 449\text{ K}, 550\text{ K}\}$. The initial weights were guessed by the SST trapezoid rule[10] (cf. Section S1, SI). An occupancy-driven weight update scheme, which is explained in ref 15 and in Section S3 of the SI, was employed to generate a uniform distribution of the replicas over the rungs during SST-MD. For each replica, the MM/PMM-SST simulation lasted 500 ps, i.e., the converged MM/PMM weights w_k were determined from 32 × 500 ps of MD trajectory.

These weights served as initial guesses for the subsequent DFT/PMM-SST-MD simulations, in which the MM model of alanine dipeptide was substituted by a DFT model. Here, we employed the BLYP[45,46] functional and the norm-conserving pseudopotentials of Troullier and Martins[47] at a plane-wave cutoff energy of 70 Ry. Note that we have shifted the energy of the DFT/PMM Hamiltonian by subtracting the energy $E_{DFT,0} = -58252$ kcal/mol of the isolated DFT solute alanine dipeptide in its minimum energy conformation. Thereby, MM/PMM and DFT/PMM potential energies are about the same, which is a prerequisite that MM/PMM weights are reasonable initial guesses for the DFT/PMM simulation. For each replica, we extended these costly hybrid MD simulations over 55 ps.

4. RESULTS AND DISCUSSION

Simulations were run on the SuperMUC Phase 2 Petascale System at the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities in Munich, employing up to 32,256 cores in a hybrid MPI/OpenMP parallel setup, which yielded about 30 ps DFT/PMM-MD trajectory per day for each of the 32 replicas. Thus, using multiple replicas in a single run makes the method highly scalable and complements the excellent scaling properties of the IPHIGENIE/CPMD interface.[35]

4.1. SST with Polarizable DOF. First, we address the size of the configurational dependence $\epsilon_k(\mathbf{R})$ (defined by eq 8) contributing to the SST scaling prefactor of the solute−solvent energy E^{ps} in eq 7. We computed the averages $\langle\ldots\rangle$ of $\epsilon_k(\mathbf{R})$ for different solvent charge scaling factors γ_k over a snapshot ensemble $\mathcal{S}_0$ of configurations $\mathbf{R} \in \mathcal{S}_0$, which comprised 60 statistically independent snapshots drawn from rung $k = 0$ of the MM/PMM-SST ensemble.

Figure 1 shows the sizes $\langle\epsilon_k\rangle/\gamma_k$ of the configuration dependent contributions in units of the dimensionless scaling factors $\gamma_k = \sqrt{T_k/T_0}$ for temperatures $T_k \in [300\text{ K}, 1000\text{ K}]$. In the MM/PMM setting (gray) the ratios $\langle\epsilon_k\rangle/\gamma_k$ decrease about linearly with γ_k, as shown by the depicted regression line. Thus,

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

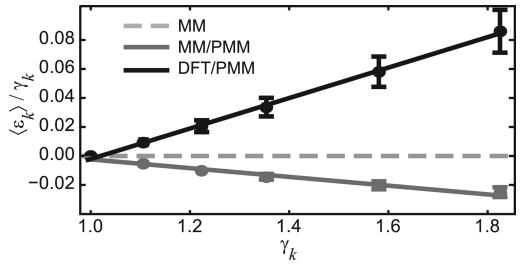


Figure 1. Relative magnitudes $\langle\epsilon_k\rangle/\gamma_k$ for MM-SST (light gray), MM/PMM-SST (gray), and DFT/PMM-SST (black) for the temperature range $T_k \in [300\text{ K}, 1000\text{ K}]$ and linear regressions (solid lines). Error bars denote standard deviations.

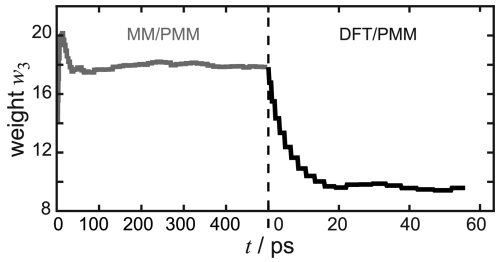


Figure 2. Evolution of the dimensionless weight w_3 of the highest rung of $\mathcal{E}$ for the MM/PMM (gray) and DFT/PMM (black) SST simulations. Note the two different scales of the time axis separated by the dashed horizontal line.

eq 8 demonstrates for MM/PMM that the absolute values of the solute−solvent interaction energies $|\langle E_k^{ps}\rangle|$ do not increase linearly with γ_k because they are always smaller than the interaction energy $|\langle E^{ps}\rangle|$ at the lowest rung after linear scaling, i.e., $|\langle E_k^{ps}\rangle| < \gamma_k\,|\langle E^{ps}\rangle|$. The two expressions would be equivalent for charge scaling in a pure MM setting (cf. eqs 2 and 3), in which $\epsilon_k(\mathbf{R})$ exactly vanishes as is indicated by the gray dashed line. For MM/PMM, the reduced effective scaling factor $\gamma_k[1+\langle\epsilon_k\rangle/\gamma_k] < \gamma_k$ of $\langle E^{ps}\rangle$ in eq 7 results from the sublinear scaling $\sim \gamma_k^{\alpha}$ $(\alpha < 1)$ of the induced PMM solvent dipoles, which is caused by the contributions of the unscaled MM solute charges to the field polarizing the solvent.

We computed $\langle\epsilon_k\rangle/\gamma_k$ also for the DFT/PMM setting (black dots and regression line, Figure 1), in which, apparently, the ratio $\langle\epsilon_k\rangle/\gamma_k$ linearly increases with γ_k. Thus, in this case, the average solute−solvent energy $|\langle E_k^{ps}\rangle|$ is always larger than the linearly scaled interaction energy $\gamma_k|\langle E^{ps}\rangle|$ because the polarization of the DFT solute is additionally enhanced by the upscaled charges of the surrounding PMM solvent.

Note, however, that the overall configurational dependence $\langle\epsilon_k\rangle/\gamma_k$ is small over the whole range of γ_k. At the highest temperature $T_3 = 550$ K ($\gamma_3 \approx 1.35$) in $\mathcal{T}$, it is below 4% and even at 1000 K ($\gamma \approx 1.83$) below 10%. No significant differences are observed, if the above analysis is repeated for a snapshot ensemble $\mathcal{S}_3$ picked from rung 3 of the PMM-SST simulation or for a ensemble $\mathcal{S}_0'$ drawn from rung 0 of the DFT/PMM-SST simulation (data not shown). Because the many solvent DOF contributing to $\epsilon_k(\mathbf{R})$ thus render only a small correction to the prefactor of E^{ps} in eq 7, the energies entering the exponents obtained from charge scaling in nonpolarizable and polarizable settings (eqs 4 and 9, respectively) are very similar. Therefore, we expect that the MM-SST concept[10] of charge scaling should yield a highly efficient sampling also for partially (MM/PMM) or fully (DFT/PMM) polarizable settings.

4.2. MM/PMM-SST and DFT/PMM-SST. Figure 2 shows the evolution of the weight w_3 resulting from the applied update scheme (cf. Section S3, SI) during MM/PMM-SST (gray) and subsequent DFT/PMM-SST (black) MD simulations. The weight w_3, which represents the free energy difference between rungs $k = 3$ and $k = 0$ (in units of $k_B T_3$), converges in MM/PMM-SST-MD quickly and reaches the value $w_3 = 17.9$ after 0.5 ns.

At the end of the MM/PMM-SST simulation, we switched to the DFT/PMM setup and equilibrated each replica on its current rung for 2 ps to dissipate excess energy due to the change of the solute description. Using the weights of the preparatory MM/PMM-SST simulation, we then continued with 55 ps DFT/PMM-SST-MD. The black line in Figure 2 reveals that w_3 is strongly diminished by about 50% to the final value $w_3 = 9.6$. This decrease is almost complete after the first 20 ps. A similar convergence behavior is observed also for w_2 and w_1 as documented by Figure S1A and B in Section S5 of the SI.

The initial overestimates of the DFT/PMM-SST weights are also witnessed in the distribution of the replicas over the rungs. For DFT/PMM-SST-MD, Figure 3A monitors the running

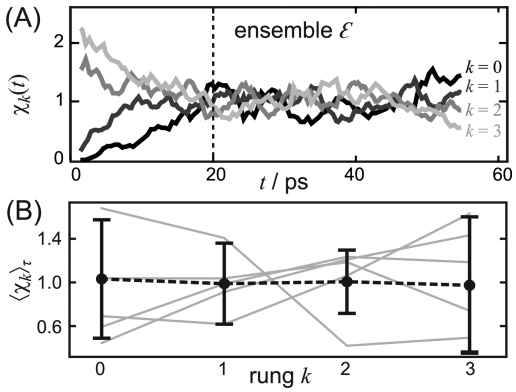


Figure 3. DFT/PMM-SST ensemble $\mathcal{E}$. (A) Running averages (window size 2.5 ps) of the uniformity measures χ_k. Only data from the time interval τ to the right of the dotted line are used for further analysis. (B) Temporal averages $\langle\chi_k\rangle_\tau$ for all (black dots) and $\langle\chi_k^c\rangle_\tau$ for selected replicas c (gray lines). The error bars denote standard deviations of χ_k.

averages of the uniformity measures[10] $\chi_k(t) = n_k(t)R/C$, which compare the number of replicas $n_k(t)$ at time t on rung k with the uniform distribution C/R characterized by $\chi_k(t) = 1\ \forall\ k$. Initially, most replicas occupy the rungs $k = 2,3$, as is witnessed by the initial values $\chi_2 \approx \chi_3 \approx 2.0$, whereas the lower rungs $k = 0,1$ are almost empty ($\chi_0 \approx \chi_1 \approx 0$). During the first 20 ps (dotted line) of weight update, the occupancies become almost uniform. In the subsequent time interval $\tau = [20\text{ ps}, 55\text{ ps}]$, all χ_k fluctuate around the mean value of 1.0 marking the uniform distribution.

Figure 3B) shows for all four rungs k the mean values $\langle\chi_k\rangle_\tau$ of the uniformity measures (black dots) together with their standard deviations (error bars) averaged over the time interval τ. All mean occupancies $\langle\chi_k\rangle_\tau$ are very close to 1.0 indicating a uniform sampling of the overall ensemble $\mathcal{E}$. For individual replicas c, this is, however, not the case during the short 35 ps trajectories as illustrated by the occupancies $\langle\chi_k^c\rangle_\tau$ of selected

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

replicas c (gray lines). These values notably differ from 1.0 explaining the sizable standard deviation of the $\langle \chi_k \rangle_\tau$. Because of the uniform sampling during the time span τ, we consider the w_k to be sufficiently converged.

With the converged weights w_k of the MM/PMM and DFT/PMM settings, the average SST exchange probabilities $\bar{p}_{ij}$ between neighboring rungs can be evaluated. Table 1 reports

Table 1. Average Exchange Probabilities $\bar{p}_{ij}$ (in %) of MM/PMM and DFT/PMM-SST Simulations of Ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$ and respective REST Simulations in $\mathcal{E}^{\mathrm{REST}}$

	$\bar{p}_{01}$	$\bar{p}_{12}$	$\bar{p}_{23}$	$\bar{\bar{p}}_{01}$	$\bar{\bar{p}}_{12}$
MM/PMM-SST	48.9	50.3	51.6	30.1	32.2
DFT/PMM-SST	48.4	47.8	50.4	27.8	30.2
MM/PMM-REST	34.4	37.7	35.2	13.2	18.3
DFT/PMM-REST	33.5	30.6	34.8	12.3	16.9

these probabilities as averages $\bar{p}_{ij} = (p_{ij} + p_{ji})/2$ over the probabilities for the upward and downward exchange direction. We have checked that within a SST simulation setting the single probabilities satisfy $p_{ij} \approx p_{ji}$ (data now shown), which is expected because our SST algorithm fulfills detailed balance and because all $\chi_k \approx 1$.

The first three SST data columns of Table 1 show that the $\bar{p}_{ij}$ are all close to 50%, which is slightly above the optimal range extending from 20% to 45%.[7] The $\bar{p}_{ij}$ of DFT/PMM are only a little smaller than those of MM/PMM. Therefore, the temperature ladder employed in the costly DFT/PMM simulations can be optimized by using the much cheaper MM/PMM approach.

The efficiency of SST is determined by the so-called round-trip rate, which is the inverse of the average time required for one replica to go from the lowest to the highest and back to the lowest rung. In the MM/PMM and DFT/PMM-SST ensembles, we observed 1223 and 82 round trips, respectively, representing similar rates of 77 and 72 ns^{-1}.

4.2.1. DFT/PMM-REST. For efficiency comparison, we additionally performed REST simulations with the charge scaling approach[10] in the MM/PMM and DFT/PMM settings. The REST ensemble employs the same temperature ladder $\mathcal{T}$ as SST. However, it features only four replicas always occupying all four rungs. Here, 940 and 180 ps of REST-MD were performed in the MM/PMM and DFT/PMM settings, respectively. The exchange probabilities $\bar{p}_{ij}$ (Table 1) show the expected result.[10] The REST values are on average about a factor of one-third smaller than those of SST. Furthermore, with REST the MM/PMM and DFT/PMM round trip rates are with 43 and 49 ns^{-1}, respectively, by 49% and 37% smaller than those of SST underlining that SST is the more efficient tempering variant.

4.3. Improved Weight Guess. Although the update dynamics quickly adapted the weights to the DFT/PMM setting, a considerable amount of computer time had to be spent on this adaptation requiring about 32 × 20 ps of DFT/PMM-MD simulation. We therefore designed a scheme to improve the initial guess for the DFT/PMM-SST weights.

Starting from weights w_k determined for a Hamiltonian H (e.g., MM/PMM), we want to estimate weights w_k' for a new Hamiltonian H' (e.g., DFT/PMM). Because the optimal weights w_k are the dimensionless free energies $\beta_k \langle E_k(\mathbf{R}) \rangle$, we need to estimate their shifts $\delta w_k = w'_k - w_k$ resulting from changing the potential energies E_k to E'_k on rung k, respectively. Assuming that the configurational space of H' can be reasonably well approximated by that of H, we calculate the required energy ensemble averages for all rungs from configurations $\mathbf{R} \in \mathcal{S}$ of a snapshot ensemble $\mathcal{S}$ drawn from the MM/PMM-SST simulation, leading to the weight shifts

$$\delta w_k = \beta_k \langle E'_k(\mathbf{R}) - E_k(\mathbf{R}) \rangle_{\mathbf{R} \in \mathcal{S}} \tag{10}$$

All corrections δw_k are then shifted by $-\delta w_0$ such that $w_0' = 0$. Equation 10 approximates the free energy perturbation (FEP) formula[48,49] in leading order (see Section S4 of the SI for a discussion). The FEP should, in principle, give more accurate results. For the setups considered here, however, it converges much slower than eq 10. Because a useful correction scheme should require only limited computational effort, we therefore restrict the weight correction to the leading term.

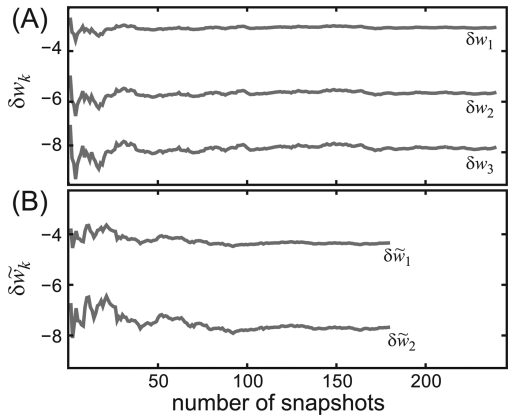


Figure 4. Convergence of the dimensionless weight corrections δw_k computed from eq 10 for increasing size of the corresponding snapshot ensemble for A) $\mathcal{E}$ and B) $\tilde{\mathcal{E}}$ ($\delta w_0 = 0$ always).

Figure 4A demonstrates that the weight corrections δw_k (eq 10) for $\mathcal{E}$ quickly converge with the size of the snapshot ensemble $\mathcal{S}$. About 50 snapshots yield converged δw_k, and only about 10 snapshots suffice for a reasonable estimate.

Note that the snapshot ensemble $\mathcal{S}$ comprises snapshots from all rungs k. Instead, one could also estimate the shifts δw_k from rung-specific ensembles $\mathcal{S}_k$, which would be physically more appropriate. The small number of snapshots, however, suffices to determine the energy averages only with a large error. These large energy fluctuations introduced by using different snapshot ensembles $\mathcal{S}_k$ were found to strongly hamper the convergence of the δw_k and thus the required number of snapshots would be impractically large.

Table 2 compares the guesses $w_k' = w_k + \delta w_k$ with the converged DFT/PMM-SST weights revealing overestimates of less than 0.5. Thus, taken as initial values of a DFT/PMM-SST simulation, these guesses should lead to a much faster weight convergence.

4.4. Improved Setup $\tilde{\mathcal{E}}$. The high SST exchange probabilities of about 50% revealed by Table 1 suggest that the number of rungs can be reduced. Therefore, we set up a second SST ensemble $\tilde{\mathcal{E}}$ with only three temperature rungs $\tilde{\mathcal{T}}$ = {300 K, 406 K, 550 K}. Here, however, we employed the weight correction scheme suggested above to compute the

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

Table 2. Weights w_k Obtained by Weight Update Dynamics in MM/PMM and DFT/PMM Settings for Ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$ (w_0 = 0 always) and MM/PMM Weights Corrected by Eq 10

	w_1	w_2	w_3	$\tilde{w}_1$	$\tilde{w}_2$
MM/PMM	8.17	14.00	17.87	11.41	17.90
$w_k + \delta w_k$	5.17	8.44	10.12	7.13	10.25
DFT/PMM	4.92	8.02	9.64	6.85	9.84

initial DFT/PMM-SST weights. The converged weights $\tilde{w}_k$ belonging to the three-rung ensemble are given in the last two columns of Table 2.

Figure 5 shows the evolution of the weight $\tilde{w}_2$ during the associated SST simulations. Here, the weight correction $\delta\tilde{w}_2$ is indicated by the light gray vertical bar, and $\tilde{w}_2$ is seen to be almost converged already at the start of the DFT/PMM-SST simulation. A similar convergence behavior is observed also for $\tilde{w}_1$ (cf. Figure S1C in Section S5 of the SI). Because the maximal temperatures T = 550 K of both ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$ are equal, also the weights at the top rung of the respective ladders (w_3 and $\tilde{w}_2$) should converge to the same value. Table 2 shows that this is actually the case.

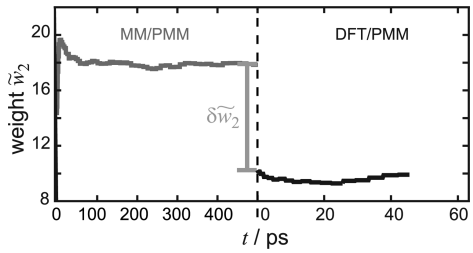


Figure 5. Evolution of the dimensionless weight $\tilde{w}_2$ of the highest rung of $\tilde{\mathcal{E}}$ for MM/PMM (gray) and DFT/PMM (black).

Besides the weights, we have monitored the temporal evolution of the four-rung ensemble $\mathcal{E}$ by the uniformity measures $\chi_k(t)$, which exhibit uniformity only after 20 ps (Figure 3A) because of strong adaption of weights in the beginning of the simulations. For the three-rung ensemble $\tilde{\mathcal{E}}$ one finds a uniform distribution right from the start of the DFT/PMM-SST simulation (see Figure S2A in Section S6 of the SI). Therefore, our new educated guess for the initial SST weights (cf. eq 10) guarantees that the replica acquire data on the target rung r = 0 straight away in the DFT/PMM-SST simulation.

The reduced number of rungs in $\tilde{\mathcal{E}}$ lowers the exchange probabilities $\bar{\tilde{p}}_{ij}$ (cf. the last two columns of Table 1) to about 30%, which now lie in the optimal range.[7] Here, the lowered $\bar{\tilde{p}}_{ij}$ are compensated by the fewer steps needed to complete a round trip on the ladder of $\tilde{\mathcal{E}}$, and concurrently, the MM/PMM- and DFT/PMM-SST round trip rates increase by about 9% to 84 and 79 ns^{-1}, respectively, indicating for the three-rung ensemble $\tilde{\mathcal{E}}$ an improved sampling efficiency at T_0.

REST simulations $\tilde{\mathcal{E}}^{\mathrm{REST}}$ with three rungs in the MM/PMM and DFT/PMM settings of 920 and 160 ps duration, respectively, revealed that the exchange probabilities are about only half as large as in the three-rung SST setups (cf. Table 1). Correspondingly, the respective round trip rates of 41 and 36 ns^{-1} and the associated sampling efficiencies are drastically smaller.

4.5. Free Energy Landscape of Alanine Dipeptide Computed by DFT/PMM-SST. Figure 6 shows the free energy landscape spanned by the two dihedral angles[50] Φ and Ψ of alanine dipeptide as predicted by our DFT model in aqueous PMM solvent. Here, we have combined data from the ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$ at T_0 = 300 K resulting from about 760 ps that have been spent on rung 0 out of a total of 2.56 ns of DFT/PMM-SST-MD trajectory accumulated by the 32 replicas. The main features of the free energy landscape are the two minima around the structural motifs "polyproline II" at $(\Phi,\Psi) \approx (-80,150)^\circ$ and "α-helical" at $(\Phi,\Psi) \approx (-90,-20)^\circ$, which are of similar depth. Conformations characterized by $\Phi > 0^\circ$ generally have a higher free energy and are separated from regions $\Phi < 0^\circ$ by high barriers but have already been visited during the SST simulations.

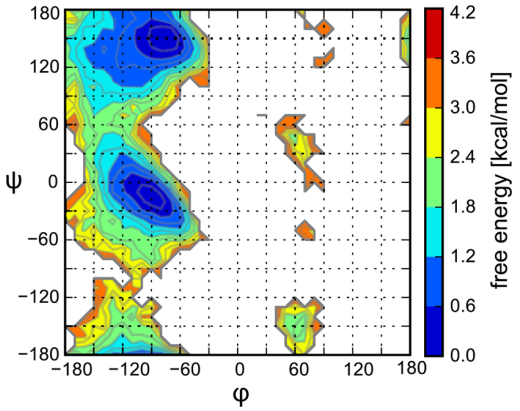


Figure 6. Current state of the free energy landscape of alanine dipeptide. DFT/PMM-SST simulation data from ensembles $\mathcal{E}$ (32 × 35 ps) and $\tilde{\mathcal{E}}$ (32 × 45 ps) have been merged; the bin size is 10° × 10°.

The overall shape of the free energy landscape has been similarly found for other QM/MM settings that employ plain[51,52] or RE[32] MD simulations using semiempirical QM Hamiltonians or a multicanonical[53] MD simulation with Hartree−Fock as the QM method. Therefore, we are confident that the two minimum energy regions are already well described by the simulations presented here. However, the relative free energies of the $\Phi > 0^\circ$ structures and the barriers connecting the various minima still carry a large error, which can be reduced only by extended simulations that are on the way. Furthermore, reweighting schemes based on the "weighted histogram analysis method" (WHAM)[54,55] or the "multistate Bennett acceptance ratio estimator" (MBAR) method[56] could be used to improve the statistics by including data from rungs $r \neq 0$. For these, however, at each trajectory sampling step the energies $E_k(\mathbf{R})$ have to be computed for *all* rungs k, which will be implemented to a future release of IPHIGENIE.

5. SUMMARY AND OUTLOOK

The conformational space of flexible molecules in polar solvents can be efficiently sampled by solute tempering methods such as REST[16] or SST.[10] This enhanced sampling concept can be transferred to simulation setups, in which the electronic polarizabilities of the solvent and, possibly, also of

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

the solute are explicitly included. Here, the key trick has been the adoption of the charge scaling method[10] to polarizable systems.

As sample systems, we have considered MM and DFT models of alanine dipeptide solvated in PMM water. The MM/PMM- and DFT/PMM-SST simulations showed rapid convergences of the required weights. The convergence of the DFT/PMM weights was significantly enhanced by deriving accurate initial weights from a preceding MM/PMM-SST simulation and very few DFT/PMM calculations. Interestingly, the enhanced sampling power of SST as compared to REST[10] fully transfers to polarizable systems. Both for SST and REST, the sampling properties of the MM/PMM and DFT/PMM setups turned out to be highly similar. Therefore, the costly DFT/PMM simulations can be optimally tuned by the much cheaper MM/PMM approach. For our DFT/PMM model of alanine dipeptide in aqueous solution and the chosen temperature range from 300 to 550 K, a three-rung ladder is sufficient for an efficient SST dynamics to sample its free energy landscape at 300 K. In contrast, conventional RE or ST simulations would require a much larger number of rungs, which would be unfeasible for the chosen level of theory (DFT). Extended simulations applying the thus developed DFT/PMM-SST-MD simulation technique and aiming at an accurate description of this landscape are under way. A highly efficient, massively parallel, and scalable DFT/PMM-SST implementation is available through the interface[30,35] of the PMM-MD program IPHIGENIE[41] with the DFT program CPMD.[42]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00951.

All technical details of the simulations (Section S1), employed PMM water model (Section S2), weight adaption scheme (Section S3), discussion of the connection between the weight correction (eq 10) and FEP formula (Section S4), time evolution of weights w_1, w_2, and $\tilde{w}_1$ in MM/PMM- and DFT/PMM-SST (Section S5), and uniformity measures for ensemble $\tilde{\mathcal{E}}$ (Section S6). (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gerald.mathias@physik.uni-muenchen.de.

Present Address

E. Gawehn: Inst. f. Pharmazeutische Wiss., ETH Zurich, Vladimir-Prelog-Weg 1−5/10, 8093 Zürich, Switzerland

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Computer time provided by the Leibniz-Rechenzentrum (PR89XE) and financial support by the Deutsche Forschungsgemeinschaft (Grant SFB749/C4) and 'Bavarian Competence Network for Technical and Scientific High Performance Computing' (KONWIHR-III) is acknowledged. We thank Prof. Paul Tavan for his critical reading of the manuscript.

■ REFERENCES

(1) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96−123.
(2) Lyubartsev, A.; Martsinovski, A.; Shevkunov, S.; Vorontsov-Velyaminov, P. *J. Chem. Phys.* **1992**, *96*, 1776−1783.
(3) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451.
(4) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(5) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(6) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(7) Lingenheil, M.; Denschlag, R.; Mathias, G.; Tavan, P. *Chem. Phys. Lett.* **2009**, *478*, 80−84.
(8) Chelli, R. *J. Chem. Theory Comput.* **2010**, *6*, 1935−1950.
(9) Denschlag, R.; Lingenheil, M.; Tavan, P. *Chem. Phys. Lett.* **2009**, *473*, 193−195.
(10) Denschlag, R.; Lingenheil, M.; Tavan, P.; Mathias, G. *J. Chem. Theory Comput.* **2009**, *5*, 2847−2857.
(11) Park, S.; Ensign, D. L.; Pande, V. S. *Phys. Rev. E* **2006**, *74*, 066703.
(12) Zhang, C.; Ma, J. *Phys. Rev. E* **2007**, *76*, 036708.
(13) Park, S.; Pande, V. S. *Phys. Rev. E* **2007**, *76*, 016703.
(14) Nguyen, P. H.; Okamoto, Y.; Derreumaux, P. *J. Chem. Phys.* **2013**, *138*, 061102.
(15) Lenner, N.; Mathias, G. *J. Chem. Theory Comput.* **2016**, DOI: 10.1021/acs.jctc.5b00751.
(16) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749−13754.
(17) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058−9067.
(18) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. *J. Chem. Theory Comput.* **2006**, *2*, 217−228.
(19) Wang, L.; Friesner, R. A.; Berne, B. *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(20) Terakawa, T.; Kameda, T.; Takada, S. *J. Comput. Chem.* **2011**, *32*, 1228−1234.
(21) Chelli, R.; Signorini, G. F. *J. Chem. Theory Comput.* **2012**, *8*, 830−842.
(22) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(23) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(24) Schwörer, M.; Mathias, G. In *Computational Trends in Solvation and Transport in Liquids*; Sutmann, G., Grotendorst, J., Gompper, G., Marx, D., Eds.; Schriften des Forschungszentrums Jülich, IAS Series *28*, 2015.
(25) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−B870.
(26) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133−A1138.
(27) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452−10467.
(28) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941−6947.
(29) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176−1184.
(30) Schwörer, M.; Breitenfeld, B.; Tröster, P.; Bauer, S.; Lorenzen, K.; Tavan, P.; Mathias, G. *J. Chem. Phys.* **2013**, *138*, 244103.
(31) Moskovsky, A.; Vanovschi, V.; Konyukhov, S.; Nemukhin, A. *Int. J. Quantum Chem.* **2006**, *106*, 2208−2213.
(32) Seabra, G. D. M.; Walker, R. C.; Roitberg, A. E. *J. Phys. Chem. A* **2009**, *113*, 11938−48.
(33) Buin, A.; Ma, J.; Huang, Y.; Consta, S.; Hui, Z. *J. Phys. Chem. C* **2012**, *116*, 8608−8618.
(34) Fedorov, D. G.; Sugita, Y.; Choi, C. H. *J. Phys. Chem. B* **2013**, *117*, 7996−8002.
(35) Schwörer, M.; Lorenzen, K.; Mathias, G.; Tavan, P. *J. Chem. Phys.* **2015**, *142*, 104108.
(36) Schwörer, M.; Wichmann, C.; Tavan, P. *J. Chem. Phys.*, submitted for publication.
(37) Lorenzen, K.; Schwörer, M.; Tröster, P.; Mates, S.; Tavan, P. *J. Chem. Theory Comput.* **2012**, *8*, 3628−3636.

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

(38) Lorenzen, K.; Wichmann, C.; Tavan, P. *J. Chem. Theory Comput.* **2014**, *10*, 3244−3259.
(39) Lorenzen, K.; Mathias, G.; Tavan, P. *J. Chem. Phys.* **2015**, *143*, 184114.
(40) Bauer, S.; Tavan, P.; Mathias, G. *J. Chem. Phys.* **2014**, *140*, 104103.
(41) IPHIGENIE (version 1.5) is available for download free of charge under the GPL licence at http://sourceforge.net/projects/iphigenie.
(42) Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. *CPMD: Car-Parinello Molecular Dynamics*, version 3.17.1.; IBM Corp 1990−2008 and MPI für Festkörperforschung Stuttgart 1997−2001. www.cpmd.org.
(43) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(44) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(45) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.
(46) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(47) Troullier, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993−2005.
(48) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(49) Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation*, 1st ed.; Oxford University Press USA: New York, 2010; Chapter 8.1, pp 314−317.
(50) Ramachandran, G.; Ramakrishnan, C. t.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95−99.
(51) Hu, H.; Elstner, M.; Hermans, J. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 451−463.
(52) Kwac, K.; Lee, K.-K.; Han, J. B.; Oh, K.-I.; Cho, M. *J. Chem. Phys.* **2008**, *128*, 105106.
(53) Jono, R.; Watanabe, Y.; Shimizu, K.; Terada, T. *J. Comput. Chem.* **2010**, *31*, 1168−1175.
(54) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(55) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 26−41.
(56) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.

DOI: 10.1021/acs.jctc.5b00951
J. Chem. Theory Comput. 2016, 12, 992−999

Der folgende Abdruck

Supporting Information for:
Simulated Solute Tempering in Fully Polarizable
Hybrid QM/MM Molecular Dynamics Simulations

Magnus Schwörer, Christoph Wichmann, Erik Gawehn und Gerald Mathias

J. Chem. Theory Comput. **12**, 992-999 (2016)

enthält zusätzliche Informationen zum oben abgedruckten Haupttext. In Abschnitt S1 werden die Details der durchgeführten Simulationen präsentiert, Abschnitt S2 dokumentiert die verwendete Variante des GP6P-Modells (die sich leicht von der in Ref. [58] entwickelten finalen Version unterscheidet). In Abschnitt S3 wird der zur Anpassung der SST-Gewichte verwendete Algorithmus diskutiert, in Abschnitt S4 werden einige zusätzliche Überlegungen zum Transfer der Gewichtsparameter w_k zwischen PMM- und DFT/PMM-SST angestellt. In den beiden letzten Abschnitten S5 und S6 werden schließlich der Vollständigkeit halber die Zeitentwicklungen von Gewichtsparametern bzw. die Gleichverteilung der Kopien k auf der Temperaturleiter T_k gezeigt.

Supporting information for: Simulated Solute Tempering in Fully Polarizable Hybrid QM/MM Molecular Dynamics Simulations

Magnus Schwörer,[†] Christoph Wichmann,[†] Erik Gawehn,[†,‡] and Gerald Mathias[*,†]

†*Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany.*

‡*Present adress: Inst. f. Pharmazeutische Wiss., ETH Zurich, Vladimir-Prelog-Weg 1-5/10, 8093 Zürich, Switzerland*

E-mail: gerald.mathias@physik.uni-muenchen.de

S1 Simulation Methods

All calculations of alanine dipeptide (Ace-Ala-NHMe) in PMM water were performed with the IPHIGENIE/CPMD program package, which combines the parallel PMM-MD code IPHIGENIE[S1–S7] with the parallel grid-based plane-wave DFT code CPMD.[S8]

In the MM/PMM setting, the dipeptide was modeled by CHARMM22/CMAP.[S9,S10] In DFT/PMM it was described by the BLYP[S11,S12] functional and the norm-conserving pseudo-potentials of Troullier and Martins[S13] at a plane-wave cutoff energy of 70 Ry. It was centered in a cubic DFT box with a volume of $(14\ \text{Å})^3$. The DFT self-consistency convergence

criterion[S4] was $\chi_{\mathrm{DFT}} = 10^{-6}$. Solute energies, which are required for the computation of the Metropolis exponents in solute tempering methods,[S14,S15] were calculated with respect to the energy $E_{\mathrm{DFT},0}$ of the isolated molecule at its relaxed geometry obtained from a single separate calculation. This energy offset $E_{\mathrm{DFT},0} = -58252\,\mathrm{kcal\,mol^{-1}}$ shifts the DFT solute energies to the same level as the MM solute energies. Otherwise, a weight shift of $\delta w_k^0 = (\beta_k - \beta_0)E_{\mathrm{DFT},0}$ – up to 44130 for w_3 – would have to be included when using the MM/PMM weights in the DFT/PMM simulation. Although this increase would be captured by the weight correction (10), the comparison of MM/PMM and DFT/PMM weights of similar magnitude is more instructive.

As PMM water model we used the Gaussian Polarizable Six-Point (GP6P) potential, which was specifically developed for DFT/PMM settings and will be presented elsewhere.[S16] Here we solely note that this PMM energy function for water was parametrized by following, extending, and partially correcting the prescriptions in Refs. [S17] and [S18]. Its parameters are documented and explained in Section S2 of this SI. The just quoted unpublished work additionally provides optimized parameters for selected Lennard-Jones (LJ) potentials of the alanine dipeptide DFT fragment embedded in GP6P water. These parameter were designed to describe the hydrogen bonding of the dipeptide's amide groups with the surrounding GP6P water in such a way that available knowledge about radial distribution functions is reproduced. The thus revised LJ parameters refer to the amide carbon (C′), oxygen (O), nitrogen (N), and hydrogen (H) atoms. They are given in Table S1. All other LJ parameters, which are required for the alanine dipeptide DFT fragment in a DFT/(P)MM setting,[S19] were adopted from the CHARMM22[S9] force field.

Table S1: LJ Parameters for amide group atoms α embedded in GP6P water.

α	A_α / (10^3 Å^{12} kcal mol^{-1})	B_α / (Å^6 kcal mol^{-1})
C′	1585.97	962.649
O	426.926	252.844
N	788.889	1357.18
H	0.0	0.0

The geometries of the water molecules were constrained with the M-SHAKE[S20] and RATTLE[S21] algorithms with relative tolerances of 10^{-10}. The equations of motion were integrated with the velocity Verlet algorithm[S22] with time steps δt of 1 fs in the MM/PMM and 0.5 fs in the DFT/PMM setting, respectively. The electrostatic and van der Waals dispersion interactions were treated by the most recent energy conserving version of the fast structure-adapted multipole method (SAMM).[S1–S3] The SAMM description of these interactions employed 4'th and 3'rd order symmetric Taylor expansions, respectively, which were extended up to a maximum distance dictated by the minimum image convention (MIC) of the applied toroidal boundary conditions.[S23] Beyond the MIC distance, a moving boundary electrostatic reaction field correction,[S24] which models a surrounding dielectric continuum with a dielectric constant of 78, and a continuum correction for the dispersion attraction were applied. The SAMM accuracy parameter Θ was set to medium ($\Theta_{\mathrm{m}} = 0.20$) in the MM/PMM and high ($\Theta_{\mathrm{a}} = 0.17$) accuracy in the DFT/PMM setting, respectively. The PMM dipole convergence criterion[S4] was set to $\chi_{\mathrm{PMM}} = 10^{-4}$ D.

A condensed phase simulation box was prepared in the MM/PMM setting by filling a periodic cubic box with one MM alanine dipeptide model (CHARMM22) and 4487 PMM water models (GP6P), comprising a total of $N = 26944$ particles (i.e. atoms and massless charge sites). The MM/PMM system was equilibrated in the NpT ensemble by a 1 ns MD simulation to the standard conditions of $T = 300$ K and $p = 1$ bar employing a Bussi thermostat[S25] (coupling time 0.1 ps) and Berendsen barostat[S26] (coupling time 10 ps), respectively. It yielded a simulation box volume of $V = (51.3\ \text{Å})^3$ corresponding to a density of $0.996\,\mathrm{g/cm^3}$. From the tail of a subsequent 1 ns MD simulation in the NVT ensemble we collected 32 snapshots at temporal distances of 5 ps.

These snapshots served as starting points for a MM/PMM-SST generalized ensemble $\mathcal{E}$ with $C = 32$ replicas and four temperature rungs $T_k \in \mathcal{T} = \{300\,\mathrm{K}, 367\,\mathrm{K}, 449\,\mathrm{K}, 550\,\mathrm{K}\}$. All subsequent SST simulations were executed in the NVT_k settings. Exchanges between rungs were attempted every $\Delta t = 250$ fs obeying the deterministic even-odd scheme.[S27,S28]

Initial guesses for the MM/PMM SST weights w_k were obtained by the following procedure. Short (10 ps) MD simulations were carried out at each temperature rung k. The resulting trajectories served to collect temporal averages $\langle E_k^{\mathrm{pp}} \rangle$ and $\langle E_k^{\mathrm{ps}} \rangle$ of the solute-solute and solute-solvent interaction energies, respectively. From these averages the weights w_k were calculated with the SST trapezoid rule {Eq. (13) in Ref. [S15]}. A subsequent MM-SST-MD simulation with a duration of 300 ps then yielded through the update rule, which will be described in Section S3, converged weights w_k^{MM}. In these preparatory simulations the water model was TIP3P.[S29]

In SST simulations, the just mentioned update rule for the weights serves to achieve a uniform distribution of replicas over the temperature rungs [cf. Eq. (S1) in Section S3]. In our applications, the initial learning parameter $\alpha(0) = 0.5$ was increased at each update step by increments $\Delta\alpha = 0.005$ until $\alpha_{\max} = 0.98$ was reached. The weights w_k were updated with a period of $n\Delta t$, where Δt is the time span between subsequent exchange attempts. The factor n was increased at each weight update by 1 % or at least 1 from its initial value $n = 2$ thus implementing a rather slow learning process. Table S2 lists the converged weights for all settings.

The MM/PMM-SST ensemble $\mathcal{E}$ was initiated with the weights w_k^{MM} and simulated for 500 ps yielding converged weights w_k. After substituting the MM solute with the DFT model a 2 ps DFT/PMM equilibration of the 32 replicas on their respective rungs was performed. Here, like in all DFT/PMM simulations, the Bussi thermostat was exclusively coupled to the solvent. Finally, the DFT/PMM-SST simulation of the four-rung ensemble $\mathcal{E}$ was started with the weights w_k.

To compute the weight corrections δw_k by Eq. (10), we drew 60 statistically independent snapshots $\mathcal{S}_k$ for each rung k from the MM/PMM trajectory and thus obtained the snapshot ensemble $\mathcal{S} = \cup_k \mathcal{S}_k$ comprising all 240 snapshots.

To test the robustness of the approach (10), which allows to accurately estimate DFT/PMM-SST weights from a few DFT/PMM calculations on MM/PMM-SST snapshots, we set

up a second generalized ensemble $\tilde{\mathcal{E}}$ featuring only three temperature rungs $\tilde{\mathcal{T}} = \{300\,\mathrm{K}, 406\,\mathrm{K}, 550\,\mathrm{K}\}$. It was prepared by the same procedures as $\mathcal{E}$. The key difference of this second DFT/PMM-SST simulation was the choice of the initial weights, which was $\tilde{w}_k + \delta\tilde{w}_k$ with corrections $\delta\tilde{w}_k$ specified by Eq. (10). The $\delta\tilde{w}_k$ were obtained from a snapshot ensemble $\tilde{\mathcal{S}}$ ($|\tilde{\mathcal{S}}| = 180$) prepared just like $\mathcal{S}$.

For a comparison of the efficiency of DFT/PMM solute tempering methods, we also set up a REST[S14] ensemble $\mathcal{E}^{\mathrm{REST}}$. It employs the same temperature ladder $\mathcal{T}$ as the SST ensemble $\mathcal{E}$. However, it features only four replicas each exclusively occupying one rung. Furthermore, we set up a three-rung REST ensemble $\tilde{\mathcal{E}}^{\mathrm{REST}}$ employing the ladder $\tilde{\mathcal{T}}$. Like in SST also in the REST simulations, the solute tempering was accomplished by charge scaling.[S15] We performed 940 ps (920 ps) of REST simulation in the MM/PMM and 180 ps (160 ps) in the DFT/PMM settings for ensemble $\mathcal{E}^{\mathrm{REST}}$ ($\tilde{\mathcal{E}}^{\mathrm{REST}}$), respectively.

Table S2: Weights w_k obtained by the SST trapezoid rule for an MM setting, and by the weight update dynamics (S1) in MM, MM/PMM and DFT/PMM settings for ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$. Note that $w_0 = \delta w_0 = 0$ always. The last line shows the MM/PMM weights w_k corrected towards DFT/PMM by Eq. (10).

	w_1	w_2	w_3	$\tilde{w}_1$	$\tilde{w}_2$
MM, SST trapezoid rule	6.40	10.44	13.14	8.45	11.84
MM-MD	6.13	10.15	12.45	8.52	12.75
MM/PMM-MD	8.17	14.00	17.87	11.41	17.90
DFT/PMM-MD	4.92	8.02	9.64	6.85	9.84
$w_k + \delta w_k$	5.17	8.44	10.12	7.13	10.25

S2 A Gaussian Polarizable Six-Point Water Model

In our MD simulations of alanine dipetide in PMM water we employed the Gaussian Polarizable Six-Point (GP6P) model potential, which features five Gaussian charges and one induced Gaussian dipole. It was developed specifically for DFT/PMM simulations. The details of the parametrization procedure and a thorough evaluation of bulk liquid phase properties will be presented elsewhere.[S16]

The GP6P water model intrinsically provides the smooth electrostatic sources required for the combination of plane-wave DFT with (P)MM force fields.[S4,S19,S30,S31] The parametrization followed the DFT/PMM-based procedure recently applied to construct polarizable water models of increasing complexity, which however employed point charges.[S17,S18] The rigid GP6P model features the experimental liquid phase geometry[S32,S33] (oxygen-hydrogen [O-H] distance $l_{\mathrm{OH}} = 0.968$ Å, HOH angle $\varphi_{\mathrm{HOH}} = 105.3°$) and the experimental vacuum dipole moment[S34] $|\boldsymbol{\mu}^{\mathrm{g}}_{\mathrm{exp}}| = 1.855$ D. The polarizability is modeled by an induced Gaussian dipole distribution of strength $\boldsymbol{\mu}^{\mathrm{i}}$ and Gaussian width $\sigma_{\boldsymbol{\mu}}$ centered at the oxygen, which isotropically and linearly depends through the experimental gas phase polarizability[S35] $\alpha = 1.47$ Å^3 on the polarizing electric field. Thus, this field is the average over a Gaussian volume of width $\sigma_{\boldsymbol{\mu}}$ around the oxygen.[S4] The width $\sigma_{\boldsymbol{\mu}}$ is specified in Table S3 together with all other parameters characterizing the employed version of the GP6P model.

Gaussian charge distributions of identical strengths q_{H} and widths σ_{H} are centered at the two hydrogen atoms. A third Gaussian charge distribution (strength q_{M}, width σ_{M}) sits at one massless site (M) located the bisectrix of φ_{HOH} at a distance l_{OM} from the oxygen towards the hydrogens. Finally, two massless lone pair sites (L) lie in the plane spanned by the bisectrix of φ_{HOH} and by the normal of the molecular plane (O-L distance l_{OL}, LOL angle φ_{LOL}, where L-sites with $\varphi_{\mathrm{LOL}} > 180°$ would exhibit perpendicular projections to positions inside the molecular triangle). Also the L-sites carry Gaussian charge distributions (strengths q_{L}, widths $\sigma_{\mathrm{L}} = \sigma_{\mathrm{M}}$).

Table S3: Parameters of the GP6P Model.

α / Å^3	1.47	l_{OH} / Å	0.968
$\sigma_{\boldsymbol{\mu}}$/ Å	0.740	l_{OM} / Å	0.555433
σ_{H} / Å	0.4588152	l_{OL} / Å	0.581108
$\sigma_{\mathrm{M/L}}$ / Å	0.6605822	φ_{HOH} / deg	105.3
q_{H} / e	0.524872	φ_{LOL} / deg	173.645
q_{M} / e	-0.449410	A_1 / (10^3 kcal mol^{-1})	12.38527
q_{L} / e	-0.300167	A_2 / Å^{-1}	2.923912
		B / (Å^6 kcal mol^{-1})	912
		A / (10^3 Å^6 kcal mol^{-1})	598.0948

The van der Waals interactions between water molecules are modeled by a single Buckingham[S36] potential $U_{\mathrm{Bu}}(r) = A_1 \exp(-rA_2) - B/r^6$ centered at the oxygen. To be capable of treating also interactions of GP6P water models with molecular models, whose van der Waals interactions are described by LJ potentials $U_{\mathrm{LJ}}(r) = A/r^{12} - B/r^6$, a LJ approximation to the GP6P Buckingham potential has been derived by requiring that, in the distance range $r \in [2.5, 6]\,\text{Å}$, the mean square deviation between $U_{\mathrm{Bu}}(r)$ and $U_{\mathrm{LJ}}(r)$ becomes minimal upon variation of A. In the current context, the approximate LJ potential has been used for the description of the van der Waals interactions between the GP6P water molecules and the MM and DFT models of alanine dipeptide. Note that the GP6P parameter list of Table S3 includes those of the approximate LJ potential.

S3 Adaptation of Weights in SST

The weights w_k entering the Metropolis exponents (1), (4) or (9) remain to be specified. We want to optimize them in such a way that all R temperature rungs k are uniformly sampled, i.e. that the probability p_k of finding the replica on rung k approaches $\bar{p} = 1/R$ during a simulation. The deviation of $p_k/\bar{p}$ from unity measures the non-uniformity of the distribution of the replica over the rungs.[S15]

Increasingly accurate estimates $\hat{p}_k(t) = N_k(t)\Delta t/t$ of the probabilities p_k are determined by the numbers $N_k(t)$ of exchange attempts starting from the rungs k and the total number $t/\Delta t$ of attempts after a SST simulation time t. Denoting the weight of a rung k at time t by $w_k(t)$, the weight update rule[S37]

$$w_k(t + n\Delta t) = w_k(t) - \ln[(1-\alpha)\chi_k(t) + \alpha] \tag{S1}$$

employs a learning parameter $0 < \alpha < 1$ to adapt the weights w_k according to the deviation of the current sampling of rung k from uniform sampling as quantified by $\chi_k(t) = \hat{p}_k(t)/\bar{p}$. The adaptation is performed every n exchange steps, i.e. with a frequency of $1/(n\Delta t)$. We

use the time-dependent learning parameter $\alpha \equiv \alpha(t) = \min[\alpha(0) + \Delta\alpha\, t/(n\Delta t), \alpha_{\max}]$ and additionally scale n by a factor $\gtrsim 1$ at each update step to gradually slow down the learning process. If a rung k is uniformly sampled at time t', i.e. $\chi_k(t') = 1$, then Eq. (S1) yields $w_k(t' + \Delta t) = w_k(t')$. In all other cases, the w_k are adjusted towards uniform sampling.

Note that all numerical values for weights are given with respect to w_0, because only weight differences enter the SST exponent (9). The scheme (S1) is easy to implement and requires only negligible computational effort.

S4 Free Energy Perturbation and Weight Correction

The free energy perturbation (FEP) equation[S38,S39]

$$\beta A_{HH'} = -\ln\langle \mathrm{e}^{-\beta(E'-E)}\rangle_{\mathcal{S}} \tag{S2}$$

estimates the free energy change $A_{HH'}$ entailed by switching from Hamiltonian H to a slightly 'perturbed' Hamiltonian H' with associated energies $E \equiv E(\mathbf{R})$ and $E' \equiv E'(\mathbf{R})$, respectively, where $\langle \ldots \rangle_{\mathcal{S}}$ denotes the average over the configurations $\mathbf{R}$ in the ensemble $\mathcal{S}$ sampled by H; we have dropped the rung index k here. Introducing the mean values $\langle E'\rangle_{\mathcal{S}}$ and $\langle E\rangle_{\mathcal{S}}$ into Eq. (S2) yields

$$\begin{aligned} \beta A_{HH'} &= -\ln\langle \mathrm{e}^{-\beta[E'-\langle E'\rangle_{\mathcal{S}}-(E-\langle E\rangle_{\mathcal{S}})+\langle E'\rangle_{\mathcal{S}}-\langle E\rangle_{\mathcal{S}}]}\rangle_{\mathcal{S}} \\ &= \beta\langle E'-E\rangle_{\mathcal{S}} - \ln\langle \mathrm{e}^{-\beta[E'-\langle E'\rangle_{\mathcal{S}}-(E-\langle E\rangle_{\mathcal{S}})]}\rangle_{\mathcal{S}} \end{aligned} \tag{S3}$$

The first term on the r.h.s. in Eq. (S3) is our approximate weight shift δw given by the correction formula (10) and captures the estimated shift of the mean values of the potential energies. The second term, which the weight correction neglects, compares the deviations $E(\mathbf{R}) - \langle E(\mathbf{R})\rangle_{\mathcal{S}}$ of the energy $E(\mathbf{R})$ of configuration $\mathbf{R}$ from the mean value $\langle E(\mathbf{R})\rangle_{\mathcal{S}}$ for the Hamiltonians H and H' over all configurations $\mathbf{R} \in \mathcal{S}$.

For our MM/PMM and DFT/PMM Hamiltonians H and H' we have calculated the exponents contributing to the ensemble average of the second term that lie in the range $[-20,\ 20]\,\mathrm{kcal\,mol^{-1}}$, i.e. the deviations of the potential energies calculated using the two solute Hamiltonians (MM or DFT) for single configurations $\mathbf{R}$ from their respective ensemble averages differ sizable. Concomitantly, we observe a very slow convergence of this second term, particularly because the exponential averaging is very sensitive to outliers. Thus, its convergence would require snapshot sets much larger than acceptable for a weight shift estimate. Here, computer time is better spent on the actual SST simulation employing H' since it allows a direct determination of the weights and one does not rely on the validity of the Zwanzig FEP formula.

S5 Time Evolution of Weights w_1, w_2, and $\tilde{w}_1$ in PMM- and DFT/PMM-SST

Figures 2 and 5 in the main text show the evolution of the weights w_3 and $\tilde{w}_2$ of the respective highest rungs of the ensembles $\mathcal{E}$ and $\tilde{\mathcal{E}}$ during the PMM- and DFT/PMM-SST simulations. Figures S1A), S1B) and S1C) document the evolution of the remaining weights w_1, w_2, and $\tilde{w}_1$, respectively.

S6 Uniformity of the Three-Rung Sampling

Figures S2A) and S2B) show the running averages $\tilde{\chi}_k(t)$ and the average uniformity measures $\langle\tilde{\chi}_k\rangle_{\tilde{\tau}}$ of the distribution χ_k^c of replicas, respectively, for the three-rung ensemble $\tilde{\mathcal{E}}$.

In contrast to the four-rung ensemble $\mathcal{E}$, for which these quantities are shown in the corresponding Figures 3A) and 3B) in the main text, the improved initial weight guesses $\tilde{w}_k + \delta\tilde{w}_k$ [cf. Eq. (10)] employed in $\tilde{\mathcal{E}}$ yield uniformity right from the start of the simulation. Accordingly, the $\langle\tilde{\chi}_k\rangle_{\tilde{\tau}}$ in S2B) are averages over the whole DFT/PMM simulation time $\tilde{\tau}$ of

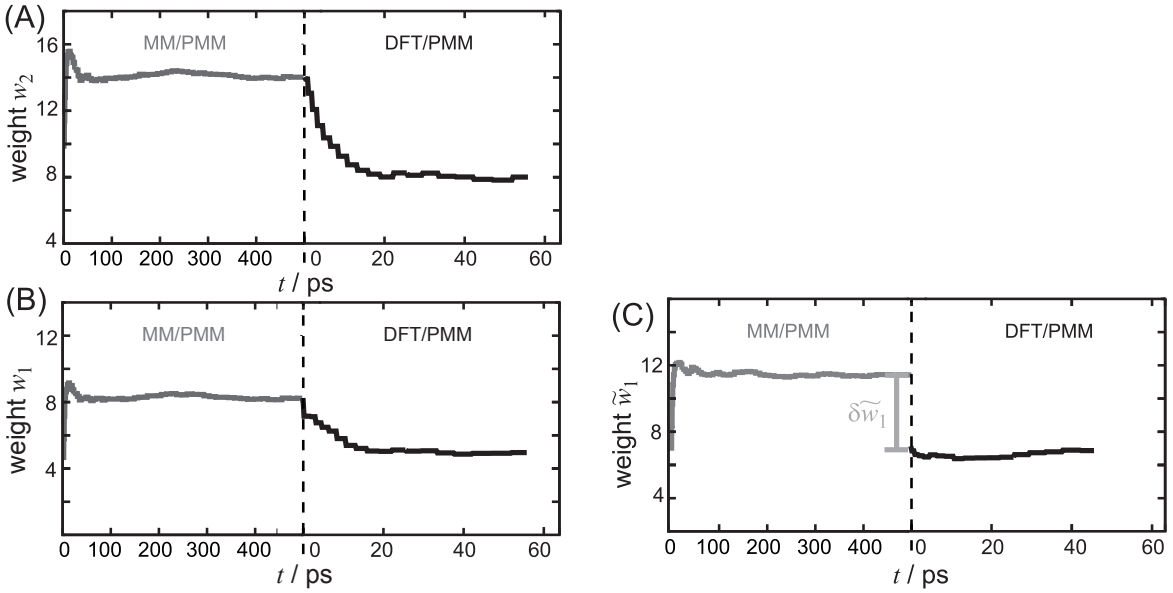


Figure S1: Evolution of the dimensionless weights A) w_2 and B) w_1 of ensemble $\mathcal{E}$ and C) $\tilde{w}_1$ of $\tilde{\mathcal{E}}$ during the MM/PMM (gray) and DFT/PMM (black) SST simulations. In panel C), $\tilde{w}_1$ has been corrected by Eq. (10) before the DFT/PMM-SST simulation. Note the two different scales of the time axis separated by the dashed horizontal line.

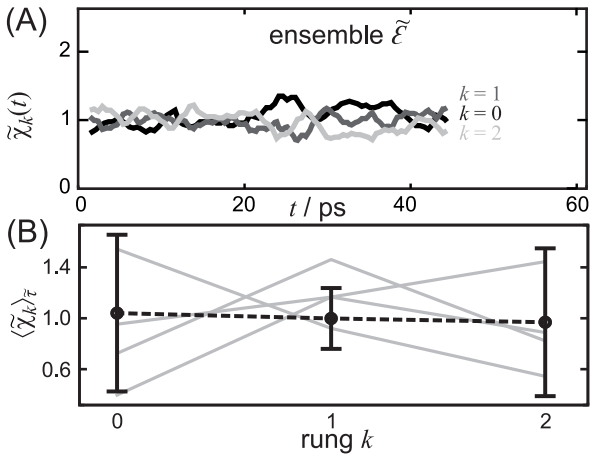


Figure S2: DFT/PMM-SST ensemble $\tilde{\mathcal{E}}$. A) Running average of the uniformity measure χ_k. B) χ_k^c for selected replicas c (gray) and the ensemble average χ_k (black). Error bars denote standard deviations.

ensemble $\tilde{\mathcal{E}}$, because no initial adaptation phase had to be discarded.

References

(S1) Lorenzen, K.; Schwörer, M.; Tröster, P.; Mates, S.; Tavan, P. *J. Chem. Theory Comput.* **2012**, *8*, 3628–3636.

(S2) Lorenzen, K.; Wichmann, C.; Tavan, P. *J. Chem. Theory Comput.* **2014**, *10*, 3244–3259.

(S3) Lorenzen, K.; Mathias, G.; Tavan, P. *J. Chem. Phys.* **2015**, *143*, 184114.

(S4) Schwörer, M.; Breitenfeld, B.; Tröster, P.; Bauer, S.; Lorenzen, K.; Tavan, P.; Mathias, G. *J. Chem. Phys.* **2013**, *138*, 244103.

(S5) Bauer, S.; Tavan, P.; Mathias, G. *J. Chem. Phys.* **2014**, *140*, 104103.

(S6) Schwörer, M.; Lorenzen, K.; Mathias, G.; Tavan, P. *J. Chem. Phys.* **2015**, *142*, 104108.

(S7) IPHIGENIE (version 1.5) is available for download free of charge under the GPL licence at `http://sourceforge.net/projects/iphigenie`.

(S8) Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. CPMD: Car–Parinello Molecular Dynamics, version 3.17.1. © IBM Corp 1990–2008 and MPI für Festkörperforschung Stuttgart 1997–2001: `www.cpmd.org`.

(S9) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(S10) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(S11) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(S12) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(S13) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2005.

(S14) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13749–13754.

(S15) Denschlag, R.; Lingenheil, M.; Tavan, P.; Mathias, G. *J. Chem. Theory Comput.* **2009**, *5*, 2847–2857.

(S16) Schwörer, M.; Wichmann, C.; Tavan, P. *submitted to J. Chem. Phys.* **2016**,

(S17) Tröster, P.; Lorenzen, K.; Schwörer, M.; Tavan, P. *J. Phys. Chem. B* **2013**, *117*, 9486–9500.

(S18) Tröster, P.; Lorenzen, K.; Tavan, P. *J. Phys. Chem. B* **2014**, *118*, 1589.

(S19) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452–10467.

(S20) Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. *J. Comput. Chem.* **2001**, *22*, 501–508.

(S21) Andersen, H. C. *J. Chem. Phys.* **1983**, *52*, 24–34.

(S22) Swope, W. C.; Anderson, H. C. *J. Chem. Phys.* **1982**, *76*, 637–649.

(S23) Allen, M. P.; Tildesley, D. *Computer Simulations of Liquids*; Clarendon: Oxford, 1987; Chapter 1.5.2-3, pp 24–29.

(S24) Mathias, G.; Egwolf, B.; Nonella, M.; Tavan, P. *J. Chem. Phys.* **2003**, *118*, 10847–10860.

(S25) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*.

(S26) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(S27) Park, S.; Pande, V. S. *Phys. Rev. E* **2007**, *76*, 016703.

(S28) Lingenheil, M.; Denschlag, R.; Mathias, G.; Tavan, P. *Chem. Phys. Lett.* **2009**, *478*, 80–84.

(S29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(S30) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941–6947.

(S31) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176–1184.

(S32) Ichikawa, K.; Kameda, Y.; Yamaguchi, T.; Wakita, H.; Misawa, M. *Mol. Phys.* **1991**, *73*, 79–86.

(S33) Thiessen, W. E.; Narten, A. H. *J. Chem. Phys.* **1982**, *77*, 2656–2662.

(S34) Clough, S.; Beers, Y.; Klein, G.; Rothman, L. *J. Chem. Phys.* **1973**, *59*, 2254–2259.

(S35) Murphy, W. F. *J. Chem. Phys.* **1977**, *67*, 5877–5882.

(S36) Buckingham, R. A.; Corner, J. *Proc. R. Soc. London, Ser. A* **1947**, *189*, 118–129.

(S37) Lenner, N.; Mathias, G. *J. Chem. Theory Comput.* **2015**, submitted.

(S38) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(S39) Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation*, 1st ed.; Oxford University Press USA: New York, 2010; Chapter 8.1, pp 314–317.

3 Resümee und Ausblick

Das Ziel dieser Arbeit war die Entwicklung einer DFT/PMM-Hybridmethode zur akkuraten und effizienten Berechnung der MD und der IR-Spektren von Biomolekülen in ausgedehnter kondensierter Phase. Ausgangspunkt war dabei die von Eichinger *et al.* vorgeschlagene DFT/MM-Technik [33]. Es sollte sowohl die Stabilität und Effizienz des Verfahrens verbessert werden, als auch die Genauigkeit der Beschreibung des MM-Fragments durch die explizite Modellierung polarisierbarer elektronischer Freiheitsgrade erhöht werden.

Die durch die von Eichinger *et al.* verwendete Partialladungsnäherung [33] verursachte Asymmetrie in der Berechnung der DFT/MM-Wechselwirkungen und die damit verbundenen Dynamikartefakte wurden erfolgreich beseitigt, indem das DFT-Fragment vollständig symmetrisch in den SAMM-Algorithmus [126] eingebettet wurde. Zusammen mit dem zusätzlich entwickelten, die Gittersymmetrie berücksichtigenden Algorithmus zur Nachführung der DFT-Box bei Translationen des DFT-Fragments, können nun Hamilton'sche DFT/PMM-MD-Simulationen von ausgedehnten Systemen durchgeführt werden [112].

Weiterhin können nun auch in Hybridrechnungen induzierbare Gauß'sche Dipole zur Modellierung der polarisierbaren Freiheitsgrade des PMM-Fragments eingesetzt werden [112]. Durch die effiziente Verzahnung der nötigen DFT- und PMM-Selbstkonsistenziterationen ist der damit verbundene Mehraufwand gegenüber DFT/MM-MD nur gering, was eine wichtige Voraussetzung für die Durchführbarkeit solcher Hybridsimulationen ist.

Die Anpassung der aktuellsten Version von SAMM [109, 126, 127] auf DFT/PMM brachte außerdem massive Steigerungen von Genauigkeit und Effizienz bei der Beschreibung der DFT/PMM-Wechselwirkungen mit sich [112, 193]. Die FMM-Entwicklungen für elektrostatische Wechselwirkungen, die in der Eichinger *et al.* zur Verfügung stehenden SAMM-Version [125] nur bis zur Ordnung $p = 2$ möglich waren, werden in der neuen SAMM-Generation bis zur Ordnung $p = 4$ ausgeführt [126]. Durch die Generalisierung des SAMM-IAC [Glg. (1.10)] [109] auf DFT/PMM, die Einführung einer neuen SAMM-Hierarchieebene im DFT-Fragment und die gewichtete Berechnung von Gyrationsradien der DFT-Ladungsverteilung konnte die Performanz (d.h. das Produkt aus Genauigkeit und Effizienz) der Auswertung von elektrostatischen DFT/PMM-Wechselwirkungen bei gegebener Ordnung $p = 4$ nochmal um etwa einen Faktor 55 gesteigert werden, was zu einer Erhöhung der Gesamtperformanz der atomaren Kraftberechnung um mehr als eine Größenordnung führte [193].

Die Implementierung von DFT/PMM wurde grundlegend technisch überarbeitet, sodass mit IPHIGENIE/CPMD nun vollständig parallelisierte, stabile Simulationen auf Höchst-

leistungsrechnern wie dem SuperMUC des Leibniz-Rechenzentrums (LRZ) möglich sind. Das Hybridprogramm IPHIGENIE/CPMD erhält dabei weitgehend die hervorragenden Skalierungseigenschaften von CPMD [152]. Die entsprechende Implementierung ist der wissenschaftlichen Öffentlichkeit zugänglich gemacht worden [193]: IPHIGENIE ist frei im Internet erhältlich [192]; die in CPMD integrierten Teile des DFT/PMM-Quellcodes sind als Patch zur Version 3.17.1 verfügbar [152] und seit Version 4.1 fester Bestandteil der CPMD-Distribution. Die Verwendung von CPMD 4.1 erforderte eine Umstellung der in CPMD integrierten Teile des DFT/PMM-Quellcodes von der Programmiersprache FORTRAN77 auf FORTRAN90 und deren Anpassung an die neuen Strukturen im Rumpfquellcode von CPMD.

Die neue DFT/PMM-Technologie wurde dann erfolgreich dazu eingesetzt, das PMM-Wassermodell GP6P zu entwickeln, das sich durch die Verwendung gaußförmiger Quellen des elektrostatischen Potentials speziell für DFT/PMM-Anwendungen eignet. Dazu wurde eine zur Parametrisierung von PMM-Punktladungsmodellen entwickelte DFT/PMM-basierte Strategie [55, 129] korrigiert und ergänzt [58].

Dieser erhebliche Zusatzaufwand war nötig, da sich das in Ref. [129] entwickelte Sechspunktmodell TL6P als fehlerhaft herausstellte: bei dessen Parametrisierung und Evaluierung [114, 129] war eine Methode zur Druckberechnung eingesetzt worden, die für die verwendete Modellklasse (polarisierbares Mehrpunktmodell) ungeeignet ist und entsprechend einige tausend Bar Mißweisung aufwies. Die zur praktischen Durchführung der DFT/PMM-gestützten Parametrisierung von Wassermodellen nötige Zusatzsoftware musste ich zusammen mit Christoph Wichmann größtenteils komplett neu entwickeln. Es stellte sich außerdem heraus, dass sich mit der verwendeten Parametrisierungsmethode [55, 129] bei Verwendung eines korrekten Barometers [14] keine hinreichend genauen Punktladungs-Sechspunktmodelle konstruieren ließen, da die von den punktförmigen lone-pair-Ladungen generierten starken elektrischen Felder eine vernünftige Parametrisierung der van der Waals-Wechselwirkungen unmöglich machten. Diese Tatsache und die geplante Verwendung in DFT/PMM-Hybridmodellen begründete die Einführung von Gaußladungen für GP6P.

Das GP6P-Modell stellte sich dann aber als Erfolg heraus. Der Vergleich der in flüssiger Phase unter Normalbedingungen (Temperatur 300 K, Druck 1 atm) berechneten Observablen (radiale Verteilungsfunktion, Dichte, isobarischer thermischer Ausdehnungskoeffizient, isothermale Kompressibilität, Wärmekapazität, Diffusionskonstante, Viskosität und Dielektrizitätskonstante) mit experimentellen Daten zeigte, dass das GP6P-Modell wichtige Eigenschaften flüssigen Wassers hinreichend genau beschreiben kann [58]. Ferner war es in der Lage, das experimentell beobachtete Temperatur-Dichte-Profil [44] mit bemerkenswerter Genauigkeit vorherzusagen [58]. Es hatte damit im Wesentlichen ähnlich günstige Eigenschaften wie das mit dem falschen Barometer fehlerhaft konstruierte und evaluierte TL6P-Modell [114, 129]. Somit steht mit GP6P nun erstmalig ein speziell für DFT/PMM-Hybridsimulationen entwickeltes, genaues polarisierbares Sechspunktmodell für Wasser zur Verfügung.

Die letzte zur Berechnung der Schwingungsspektren von (Poly-)Peptiden nötige Vorarbeit war schließlich die Bestimmung der Lennard-Jones-Potentiale zur Modellierung

der van der Waals-Wechselwirkungen zwischen den DFT-Atomen einer AG und den GP6P-Molekülen. Durch die Entwicklung eines effizienten DFT/PMM-mean-field-Ansatzes konnte die Parametersuche weitgehend automatisiert werden. Die gefundenen Parameter resultierten in einer sehr guten Beschreibung der radialen Nahstruktur der PMM-Moleküle um das DFT-Fragment. Die Verwendung der CHARM22-Parameter [66] führte hingegen zu einer signifikanten Überstrukturierung der Wasserumgebung von NMA.

Schließlich wurden für drei verschiedene DFT-Modelle (MT/BP, MT/BLYP und MT/B3LYP) von NMA dessen Schwingungsspektrum in der Gasphase und in durch GP6P beschriebener wässriger Lösung mittels FTTCF aus ausgedehnten DFT-MD- bzw. DFT/PMM-MD-Trajektorien berechnet. Nach einer globalen Frequenzskalierung stimmten die MT/BLYP- und besonders die MT/BP-Gasphasenspektren sehr gut mit der experimentellen Referenz [10] überein. Für das erheblich rechenaufwändigere MT/B3LYP-Modell war diese Übereinstimmung bereits ohne Frequenzskalierung sehr gut [58]. Auch in wässriger Lösung lieferte das mittels (MT/B3LYP)/PMM-MD berechnete Spektrum eine sehr gute Vorhersage für die experimentelle Beobachtung. Das weniger rechenaufwändige (MT/BP)/PMM-Modell war nach Frequenzskalierung ebenfalls sehr gut mit dem Experiment vergleichbar und konnte auch Isotopeneffekte mit hoher Genauigkeit vorhersagen [58].

Das DFT/PMM-Hybridmodell kann also die Solvatisierungseffekte der wässrigen PMM-Umgebung auf das DFT-Fragment mit hoher Genauigkeit beschreiben. Die Fehler des mit einem einfacheren (MT/BP)/MM-Modell berechneten Spektrums von NMA [63] treten beim elaborierteren (MT/BP)/PMM-Modell nicht mehr auf [58]. Eine einzelne 70 ps lange (MT/BP)/PMM-MD Trajektorie von NMA in GP6P Wasser konnte innerhalb eines Tages berechnet werden, eine Aufgabe, die in früheren Arbeiten [63] noch ein halbes Jahr in Anspruch genommen hat. Dieser Fortschritt ist zum einen auf die in dieser Arbeit erreichten Effizienzsteigerungen zurückzuführen, zum anderen auf die ermöglichte Nutzbarmachung von Höchstleistungsrechnern wie dem vom LRZ betriebenen SuperMUC (und natürlich auf die Verfügbarkeit entsprechender Rechenzeitressourcen).

Die Kombination von DFT/PMM-MD mit dem generalisierten Ensembleverfahren SST [195] erlaubt nun außerdem die effiziente Abtastung der Konformationslandschaften von Polypeptiden (vgl. Abschnitt 1.1.3), wie am Beispiel von Alanindipeptid gezeigt wurde [194]. Diese Effizienzsteigerung ist eine wichtige Voraussetzung für das Generieren von geeigneten Startensembles für die anschließende Untersuchung der konformationsspezifischen Schwingungsspektren von Polypeptiden.

Durch die im Rahmen dieser Dissertation erreichten algorithmischen und technischen Fortschritte ist so ein öffentlich zugängliches, hoch leistungsfähiges Programmpaket entstanden. Zusammen mit dem entwickelten DFT/PMM-Modell für AGn in wässriger Lösung steht nun ein genauer und effizienter theoretischer Zugang zu den IR-Spektren und Konformationslandschaften größerer Polypeptide und Proteine zur Verfügung.

Im Rest dieses Kapitels wird nun ein Ausblick auf zukünftige Weiterentwicklungen und Anwendungen der DFT/PMM-Methode gegeben.

Anpassung von SPLAM an DFT/PMM

Bei der Entwicklung der DFT/PMM-Methode standen chemisch ungebundene DFT- und PMM-Fragmente im Vordergrund. Zur Anwendung des Verfahrens auf Systeme, in denen chemische Bindungen zwischen den Fragmenten existieren (wie z.B. im Falle des kovalent an das Protein Bacteriorhodopsin gebundenen Farbstoffs Retinal [36]), muss eine Link-Atom-Methode wie das von Eichinger *et al.* im Rahmen von DFT/MM entwickelte SPLAM-Verfahren verwendet werden [33, 176]. Bei dieser Link-Atom-Methode wird in eine C_q–C_m-Einfachbindung zwischen zwei Kohlenstoffen C_q und C_m des DFT- bzw. PMM-Fragments ein zusätzliches H-Atom — das Link-Atom L — ins DFT-Fragment gesetzt, das die Valenz von C_q absättigt. Der Einfluß dieses zusätzlichen Atoms wird dann durch verschiedene Korrekturverfahren so weit wie möglich wieder entfernt.

Das SPLAM-Konzept umfasst vier wesentliche Punkte, deren Übertragbarkeit ins neue Hamilton'sche DFT/PMM-Schema im Folgenden diskutiert wird:

(i) Setzen des Atoms L auf die C_q–C_m-Verbindungslinie in einem Abstand r_{CH} vom C_q-Atom, der aus den Gleichgewichtslagen und Kraftkonstanten von harmonisch genäherten C–C- und C–H-Bindungen bestimmt wird [Glg. (10) in Ref. [33]],

(ii) Umverteilung der auf Atom L wirkenden Kräfte auf C_q und C_m, Einführung entsprechender Energiekorrekturen $\Delta E_{\text{stretch}}$ bzw. ΔE_{angle} [Glgn. (11) und (12) in Ref. [33]] sowie Berücksichtigen von MM-Winkel- und MM-Diederpotentialen zwischen den Fragmenten [Glg. (13) in Ref. [33]],

(iii) Generalisieren des Konzeptes der 1-M-exclusion (d.h. des Abschaltens langreichweitiger Wechselwirkungen für nahe chemisch gebundene Atome, vgl. Abschnitt 1.2.1) durch Skalieren des von C_m und dessen nahen PMM-Bindungspartnern auf dem DFT-Gitter erzeugten elektrostatischen Potentials mittels einer glatten abstandsabhängigen Schaltfunktion $\tilde{\epsilon}(r)$ [siehe S. 10459 in Ref. [33]] und

(iv) Korrektur der elektrostatischen Störung, die durch das Ersetzen der unpolaren C_q–C_m-Bindung mit der polaren C_q–H-Bindung eingeführt wurde, sowie Korrektur der mit L eingeführten zusätzlichen van der Waals-Wechselwirkungen innerhalb des DFT-Fragments.

Die Punkte (i) und (ii) sind unverändert auf DFT/PMM übertragbar. Die mit dem Setzen des Link-Atoms und der Kraftumverteilung verbundenen Energiekorrekturen $\Delta E_{\text{stretch}}$ bzw. ΔE_{angle} erfordern allerdings eine Näherung der durch DFT beschriebenen C–H-Bindung durch ein harmonisches Potential, die die Hamilton'sche DFT/PMM-Dynamik stört. Diese Störung sollte jedoch klein sein.

Das Ausblenden von Wechselwirkungen (iii) kann ebenfalls leicht in die neue DFT/PMM-Methode übernommen werden. Die von Eichinger *et al.* vorgeschlagene Schaltfunktion $\tilde{\epsilon}(r)$ blendet das von C_m und dessen nahen PMM-Bindungspartnern erzeugte elektrostatische Potential unterhalb einer Entfernung von 3 Å zwischen PMM-Atom und DFT-Gitterpunkt über eine Länge von 2,5 Å glatt aus [33]. Da das Voxel-IAC [Glg. (5) in Ref. [193]] bei typischen Voxelgrößen von $R_\Lambda \approx 0{,}35$ Å eine FMM-Approximation erst

ab einem Atom-Gitterpunkt-Abstand von etwa 4 Å erlaubt, wirkt die Potentialskalierung also ausschließlich auf der Ebene der direkten Wechselwirkungen zwischen PMM-Atomen und Gitterpunkten [193]. Damit kann die Skalierung ohne großen Aufwand in die Potentialberechnung implementiert werden. Auch die symmetrische Behandlung der Skalierung bei der Berechnung der Rückwirkung der DFT-Atome auf das PMM-Fragment stellt keine Schwierigkeit dar; der Hamilon'sche Charakter der DFT/PMM-Kopplung wird also nicht gestört.

Die größte Schwierigkeit stellt Punkt (iv) dar, d.h. die Beseitigung des durch das Link-Atom eingeführten Dipolmoments $\boldsymbol{\mu}_L$ der C_q–H-Bindung. Eichinger *et al.* greifen hier, wie schon bei der Berechnung der Rückwechselwirkung des DFT-Fragments auf die MM-Atome, auf die Partialladungsnäherung zurück, und verschieben die Ladung q_L des Atoms L einfach an den Ort von C_q [33]. So kann die Energie von $\boldsymbol{\mu}_L$ im äußeren Feld durch einen einfach Term $\Delta E_{\text{dip,ext}}$ [Glg. (14) in Ref. [33]] korrigiert und die Rückwechselwirkung aus den korrigierten Partialladungen berechnet werden. Analog wird die Korrektur $\Delta E_{\text{dip,int}}$ [Glg. (15) in Ref. [33]] der Energie der übrigen DFT-Atome im Potential des Dipols $\boldsymbol{\mu}_L$ behandelt. Die entsprechenden Kraftkorrekturen sind durch einfache Gradientenbildung zugänglich [33].

Ein exaktes Eliminieren von $\boldsymbol{\mu}_L$ ist auch im neuen DFT/PMM-Verfahren unmöglich. Eine erste Näherung wäre, die von Eichinger *et al.* vorgeschlagene partialladungsbasierte Korrekturmethode auch auf DFT/PMM anzuwenden. Eine energieerhaltende Berechnung der Partialladungen [204] könnte dabei die Störung des Hamilton'schen Charakters von DFT/PMM klein halten.

Die Verschiebung der mit L assoziierten Ladungsverteilung ρ_L (im Falle von Eichinger *et al.* die Partialladung q_L) kann im DFT/PMM-Verfahren auch über die dem Atom L zugeordneten Voxel V_l erfolgen. Dabei würden, nach Berechnung der DFT-Elektronendichte, alle Gitterpunkte der Voxel V_l (und damit ρ_L) an einer senkrecht zum Verbindungsvektor $\mathbf{r}_{LC_q}$ zwischen L und C_q stehenden Schnittebene gespiegelt werden. Da die Spiegelungsoperation im Allgemeinen nicht die Gittersymmetrie erhält, müssen die Ladungen der gespiegelten Gitterpunkte anschließend auf die Punkte des regulären Gitters verteilt werden. Der Korrekturterm $\Delta E_{\text{dip,ext}}$ kann dann z.B. als die Energie der Multipolmomente von ρ_L im vom PMM-Fragment erzeugten elektrostatischen Potential Φ genähert werden. Die Rückwechselwirkung auf die PMM-Atome wird aus der gespiegelten Ladungsdichte ρ_L berechnet. Eine Korrektur $\Delta E_{\text{dip,int}}$ innerhalb des DFT-Fragments kann durch explizite Berechnung der Energie der übrigen DFT-Atome im von ρ_L vor bzw. nach der Spiegelung erzeugten Potential erfolgen.

Beide skizzierten Alternativen zur Eliminierung von $\boldsymbol{\mu}_L$ sind mit erhöhtem Rechenaufwand verbunden. Außerdem ist die Störung der Hamilton'schen DFT/PMM-Dynamik bzw. das eventuelle Auftreten weiterer Artefakte in beiden Fällen nur schwer abschätzbar. Weiterhin beruht die Zuordnung der Voxel V_l zum Atom L auf rein geometrischen Überlegungen und ist daher nicht notwendigerweise physikalisch sinnvoll.

Die Korrektur der mit Atom L zusätzlich eingeführten van der Waals-Wechselwirkungen mit den übrigen DFT-Atomen kann, wie die Punkte (i) und (ii), ebenfalls problemlos aus Ref. [33] in DFT/PMM übernommen werden. Die Approximation der DFT-Wech-

selwirkungen durch ein Lennard-Jones-Potential ist allerdings suboptimal. Zum Beispiel sollte der Dispersionsanteil nur dann korrigiert werden, wenn die verwendete DFT-Methode die Dispersionswechselwirkung tatsächlich erfassen kann [205], z.B. indem eine entsprechende empirische Korrektur angewendet wird [206].

Kopplung mit einem Spektroskopischen Kraftfeld

Wie die DFT/MM-Studien der IR-Spektren von Flavin- und Retinalchromophoren in ihrer nativen Proteinumgebung gezeigt haben, ist für eine akkurate Berechnung dieser Spektren eine korrekte Beschreibung der Polarisationseffekte im Protein notwendig [36, 37]. Das im Rahmen der laufenden Dissertation von Christoph Wichmann weiterentwickelte spektroskopische polarisierbare *sPMM*-Kraftfeld [63, 94] für AGn kann diese Effekte, im Gegensatz zu einem herkömmlichen MM-Proteinkraftfeld, erfassen.

Durch eine Kombination aus polarisierbaren Kraftkonstanten, einem Dipolflussmodell und einem fluctuating-charge-Modell können die Polarisationseffekte auf die Schwingungsspektren und auf die Ladungsverteilung der AGn modelliert, und eine Hamilton'sche Dynamik berechnet werden. So kann eine genauere Beschreibung der Proteinumgebung eines DFT-Fragments erfolgen. Eine solche DFT/(s)PMM-Kopplung würde weiterhin die Verwendung eines um polarisierbare Kraftkonstanten erweiterten flexiblen PMM-Wassermodells in Hybridsystemen erlauben.

Um das sPMM-Kraftfeld zur Beschreibung der Umgebung eines DFT-Fragments einzusetzen, sind noch einige, weitgehend technische Vorarbeiten nötig. Diese umfassen die Berücksichtigung der in sPMM zusätzlich eingeführten elektrostatischen Quellen bei der Berechnung der Wechselwirkungen mit dem DFT-Fragment, sowie die Entwicklung eines effizienten Schemas zum schnellen Erreichen der gleichzeitigen Selbstkonsistenz zwischen DFT-Fragment und den polarisierbaren sPMM-Freiheitsgraden. Da das sPMM-Kraftfeld teilweise Punktladungen verwendet müsste außerdem die für die Glättung des Potentials notwendigen Skalenparameter festgelegt werden. Auch die Potentialparameter der Lennard-Jones-Wechselwirkung zwischen DFT- und sPMM-Fragment sind noch unbekannt.

Anbindung alternativer DFT-Programme

Die aktuelle Implementierung der DFT/PMM-Methode verwendet das gitterbasierte DFT-Programm CPMD [152], das ebene Wellen als Basisfunktionen zur Darstellung der Kohn-Sham-Orbitale benutzt (siehe Abschnitt 1.2.2). Die Berechnung des Hartree-Fock-Austauschs im B3LYP-Hybridfunktional [146, 147] ist in solchen Implementierungen allerdings sehr rechenaufwändig. Die in Ref. [58] gezeigte hohe Qualität der mit (MT/B3LYP)/PMM-MD berechneten Spektren, die eine Frequenzskalierung unnötig macht, ist daher nur für kleine DFT-Fragmente wie dem NMA zugänglich. Für größere Moleküle muss auf ein einfacheres Funktional wie BP zurückgegriffen werden.

Abhilfe würde hier eine Kopplung zu einer effizienteren DFT-Implementierung wie dem QUICKSTEP-Modul [154] im Programmpaket CP2K [155] schaffen. Da die für entwickelten DFT/PMM-Algorithmen keinerlei CPMD-spezifischen Annahmen enthalten,

ist die Methode grundsätzlich mit jedem gitterbasierten DFT-Programm kompatibel. Für eine konkrete Implementierung müssten entweder die CPMD-seitigen Codeteile der DFT/PMM-Kopplung in den Quellcode von QUICKSTEP überführt werden. Alternativ könnte die Potentialberechnung auf dem Gitter vollständig auf IPHIGENIE übertragen werden, was die nötigen Änderungen am DFT-Quellcode minimieren und damit die Anbindung weiterer gitterbasierter DFT-Programme erheblich erleichtern würde.

Vollständige Evaluierung von GP6P

Das GP6P-Wassermodell wurde speziell als Lösungsmittel für DFT-Fragmente unter Normalbedingungen entwickelt; entsprechend wurde auch nur eine begrenzte Anzahl von Observablen in der flüssigen Phase bestimmt. Die Tatsache, dass das berechnete Temperatur-Dichte-Profil die experimentelle Evidenz hervorragend vorhersagt, weckt die Vermutung, dass das GP6P-Modell auch weitere Eigenschaften von Wasser in anderen Phasen und/oder unter anderen thermodynamischen Bedingungen, wie z.B. die Schmelztemperatur oder das Verhalten an Gas-Wasser-Grenzflächen mit hoher Qualität beschreiben kann. Eine solche vollständige Evaluierung eines Wassermodells ist eine Routineaufgabe [86], und könnte die Frage beantworten, in welchem Maße GP6P zur theoretischen Untersuchung der Eigenschaften reinen Wassers geeignet ist.

Erweiterung der durch DFT/PMM beschreibbaren Molekülklassen

Für die Anwendung von DFT/PMM-MD auf andere Molekülklassen als (Poly-)Peptiden kann die in Ref. [58] (Abschnitt 2.3) am Beispiel von AGn beschriebene Parametrisierung der Lennard-Jones-Potentiale zwischen dem DFT- und dem PMM-Fragment als Blaupause dienen. Die zur Optimierung der Lennard-Jones-Parameter nötigen DFT-MD-Referenzsimulationen kleiner, repräsentativer molekularer Strukturmotive sind mittlerweile mit vertretbarem Rechenaufwand möglich. Die DFT/PMM-mean-field-Methode erlaubt dann eine effiziente Optimierung von Parametern.

Vor allem die Phosphationen, deren IR-Spektren durch DFT/MM nur ungenügend beschrieben wurden [35], bieten sich für die Entwicklung weiterer DFT/PMM-Modelle an. Erstens wurde die entsprechende DFT-MD-Referenzrechnung bereits durchgeführt [164], zweitens stellt die korrekte Beschreibung der Schwingungsspektren der geladenen Phosphate offenbar ein herausforderndes Testszenario für eine Hybridmethode dar [35]. Schließlich wäre ein solches Modell für Phosphationen ein erster Schritt zur DFT/PMM-Beschreibung der Schwingungsspektren der Bausteine der DNA, der Nukleinsäuren.

Die in Ref. [58] ausführlich dokumentierte und korrigierte Parametrisierung [55, 129] des GP6P-Modells für Wasser kann außerdem als Leitfaden für die DFT/PMM-basierte Entwicklung weiterer komplexer PMM-Lösungsmittelmodelle zur Verwendung in DFT/PMM-Simulationen dienen.

Anwendungen

Mit neuen DFT/PMM-Technologie zur Beschreibung von Lösungsmitteleffekten auf die IR-Spektren von Biomolekülen und dem zur Verfügung stehenden DFT/PMM-Modell für AGn in Wasser können nun weiterführende Untersuchungen an größeren Polypeptiden in Angriff genommen werden.

Das DFT/PMM-SST-Verfahren [194] ermöglicht den Zugang zu Konformationslandschaften von Molekülen wie z.B. Alanindipeptid. Anschließend können konformationsspezifische IR-Spektren berechnet werden. Durch die hohe Effizienz der Methode kann die für statistisch fundierte Aussagen bzw. GNC-Analysen [185, 186] nötige große Datenbasis generiert werden. Eine entsprechende Veröffentlichung ist in Vorbereitung.

Das größte im Rahmen dieser Arbeit untersuchte DFT-Fragment, Alanindipeptid, stellt mit 22 Atomen noch keine besondere Herausforderung für die Stabilität und Skalierbarkeit der DFT/PMM-Methode dar. Erste Testrechnungen an einem größeren System, einem in Wasser gelösten *Tryptophan zipper*-Motiv [207] (220 DFT-Atome, 14503 GP6P-Wassermoleküle) wurden bereits erfolgreich von Gerald Mathias auf dem SuperMUC durchgeführt. Nach einigen notwendigen programmiertechnischen Änderungen, die u.a. der Verringerung des Speicherbedarfs dienten, waren stabile DFT/PMM-MD-Simulationen auch für dieses vergleichsweise große Hybridsystem ohne algorithmische Probleme möglich.

Abbildungsverzeichnis

Literaturverzeichnis

[1] Berg, J. M., J. L. Tymoczko und L. Stryer. *Biochemie*. Springer, Heidelberg - Berlin, 7. Auflage (2014).

[2] Schulz, G. E. und R. H. Schirmer. *Principles of Protein Structure*. Springer-Verlag, New York (1979).

[3] McCammon, J. A. und S. C. Harvey. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge (1987).

[4] Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

[5] Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).

[6] Goedert, M. Alpha-synuclein and neurodegenerative diseases. *Nat. Rev. Neurosci.* **2**, 492–501 (2001).

[7] Hardy, J. und D. J. Selkoe. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).

[8] Taylor, J. P., J. Hardy und K. H. Fischbeck. Toxic proteins in neurodegenerative disease. *Science* **296**, 1991–1995 (2002).

[9] Maréchal, Y. *The hydrogen bond and the water molecule: The physics and chemistry of water, aqueous and bio-media*. Elsevier (2006).

[10] Kubelka, J. und T. A. Keiderling. Ab Initio Calculation of Amide Carbonyl Stretch Vibrational Frequencies in Solution with Modified Basis Sets. 1. N-Methyl Acetamide. *J. Phys. Chem. A* **105**, 10922–10928 (2001).

[11] Serdyuk, I., N. Zaccai und G. Zaccai. *Methods in molecular biophysics: structure, dynamics, function*. Cambridge University Press (2007).

[12] Alder, B. J. und T. E. Wainwright. Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **27**, 1208–1209 (1957).

[13] van Gunsteren, W. F. und H. J. C. Berendsen. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023 (1990).

[14] Allen, M. P. und D. Tildesley. *Computer Simulations of Liquids*. Clarendon, Oxford (1987).

[15] Karplus, M. und J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **9**, 646–652 (2002).

[16] Brünger, A. T., J. Kuriyan und M. Karplus. Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460 (1987).

[17] Nilsson, L., G. M. Clore, A. M. Gronenborn, A. T. Brünger und M. Karplus. Structure refinement of oligonucleotides by molecular dynamics with nuclear Overhauser effect interproton distance restraints: application to 5' d (CGTACG) 2. *J. Mol. Biol.* **188**, 455–475 (1986).

[18] Grubmüller, H., B. Heymann und P. Tavan. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997–999 (1996).

[19] Spörlein, S., H. Carstens, H. Satzger, C. Renner, R. Behrendt, L. Moroder, P. Tavan, W. Zinth und J. Wachtveitl. Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA* **99**, 7998–8002 (2002).

[20] Mayor, U., N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. Freund, D. O. Alonso, V. Daggett und A. R. Fersht. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863–867 (2003).

[21] Schrader, T. E., W. J. Schreier, T. Cordes, F. O. Koller, G. Babitzki, R. Denschlag, C. Renner, M. Löweneck, S.-L. Dong, L. Moroder, P. Tavan und W. Zinth. Light Triggered β-Hairpin Folding and Unfolding. *Proc. Natl. Acad. Sci. USA* **104**, 15729–15734 (2007).

[22] Denschlag, R., W. J. Schreier, B. Rieff, T. E. Schrader, F. O. Koller, L. Moroder, W. Zinth und P. Tavan. Relaxation time prediction for a light switchable peptide by molecular dynamics. *Phys. Chem. Chem. Phys.* **12**, 6204–6218 (2010).

[23] Nonella, M. Structures and vibrational spectra of p-benzoquinone in different oxidation and protonation states: A density functional study. *J. Phys. Chem. B* **101**, 1235–1246 (1997).

[24] Schmitz, M., P. Tavan und M. Nonella. Vibrational analysis of carbonyl modes in different stages of light-induced cyclopyrimidine dimer repair reactions. *Chem. Phys. Lett.* **349**, 342–348 (2001).

[25] Rieff, B., S. Bauer, G. Mathias und P. Tavan. IR spectra of flavins in solution: DFT/MM description of redox effects. *J. Phys. Chem. B* **115**, 2117–2123 (2011).

[26] Barth, A. und C. Zscherp. What vibrations tell about proteins. *Q. Rev. Biophys.* **35**, 369–430 (2002).

[27] Schmitz, M. und P. Tavan. On the Art of Computing the IR Spectra of Molecules in Condensed Phase. In E. Starikov, S. Tanaka und J. Lewis (Herausgeber), *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, Kap. 8, S. 157–177. Elsevier, Amsterdam (2006).

[28] Orozco, M. und F. J. Luque. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **100**, 4187–4226 (2000).

[29] Tavan, P., H. Carstens und G. Mathias. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In J. Buchner und T. Kiefhaber (Herausgeber), *Protein Folding Handbook*, Band 1, S. 1170–1195. Wiley-VCH, Weinheim (2005).

[30] Ren, P., J. Chun, D. G. Thomas, M. J. Schnieders, M. Marucho, J. Zhang und N. A. Baker. Biomolecular electrostatics and solvation: a computational perspective. *Q. Rev. Biophys.* **45**, 427–491 (2012).

[31] Warshel, A. und M. Levitt. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).

[32] Nobelprize.org. Press Release: The 2013 Nobel Prize in Chemistry. `http://nobelprize.org/nobel_prizes/chemistry/laureates/2013/press.html`, besucht am 9. Februar 2016.

[33] Eichinger, M., P. Tavan, J. Hutter und M. Parrinello. A Hybrid Method for Solutes in Complex Solvents: Density Functional Theory Combined with Empirical Force Fields. *J. Chem. Phys.* **110**, 10452–10467 (1999).

[34] Nonella, M., G. Mathias, M. Eichinger und P. Tavan. Structures and vibrational frequencies of the quinones in Rb. Sphaeroides derived by a combined density functional / molecular mechanics approach. *J. Phys. Chem. B* **107**, 316–322 (2003).

[35] Klähn, M., G. Mathias, C. Kötting, J. Schlitter, M. Nonella, K. Gerwert und P. Tavan. IR spectra of phosphate ions in aqueous solution: Predictions of a DFT/MM approach compared with observations. *J. Phys. Chem. A* **108**, 6186–6194 (2004).

[36] Babitzki, G., G. Mathias und P. Tavan. The Infrared Spectra of the Retinal Chromophore in Bacteriorhodopsin Calculated by a DFT/MM Approach. *J. Phys. Chem. B* **113**, 10496–10508 (2009).

[37] Rieff, B., S. Bauer, G. Mathias und P. Tavan. DFT/MM description of flavin IR spectra in BLUF domains. *J. Phys. Chem. B* **115**, 11239–11253 (2011).

[38] Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).

[39] Dill, K. A. und J. L. MacCallum. The protein-folding problem, 50 years on. *Science* **338**, 1042–1046 (2012).

[40] Ball, P. *H_2O: Biographie des Wassers*. Piper, München, 2. Auflage (2002).

[41] Eisenberg, D. S. und W. Kauzmann. *The structure and properties of water*. Clarendon Press Oxford (1969).

[42] Finney, J. L. Water? What's so special about it? *Philos. Trans. R. Soc., B* **359**, 1145–1165 (2004).

[43] Kaatze, U. Complex Permittivity of Water as a Function of Frequency and Temperature. *J. Chem. Eng. Data* **34**, 371–374 (1989).

[44] Kell, G. Precise Representation of Volume Properties of Water at One Atmosphere. *J. Chem. Eng. Data* **12**, 66–69 (1967).

[45] Benedict, W. S., N. Gailar und E. K. Plyler. Rotation-Vibration Spectra of Deuterated Water Vapor. *J. Chem. Phys.* **24**, 1139–1165 (1956).

[46] Clough, S., Y. Beers, G. Klein und L. Rothman. Dipole Moment of Water from Stark Measurements of H_2O, HDO, and D_2O. *J. Chem. Phys.* **59**, 2254–2259 (1973).

[47] Verhoeven, J. und A. Dymanus. Magnetic Properties and Molecular Quadrupole Tensor of the Water Molecule by Beam-Maser Zeeman Spectroscopy. *J. Chem. Phys.* **52**, 3222–3233 (1970).

[48] Ichikawa, K., Y. Kameda, T. Yamaguchi, H. Wakita und M. Misawa. Neutron-Diffraction Investigation of the Intramolecular Structure of a Water Molecule in the Liquid Phase at High Temperatures. *Mol. Phys.* **73**, 79–86 (1991).

[49] Thiessen, W. E. und A. H. Narten. Neutron Diffraction Study of Light and Heavy Water Mixtures at 25 C. *J. Chem. Phys.* **77**, 2656–2662 (1982).

[50] Schropp, B. und P. Tavan. Flexibility does not change the polarizability of water molecules in the liquid. *J. Phys. Chem. B* **114**, 2051–2057 (2010).

[51] Murphy, W. F. The Rayleigh Depolarization Ratio and Rotational Raman Spectrum of Water Vapor and the Polarizability Components for the Water Molecule. *J. Chem. Phys.* **67**, 5877–5882 (1977).

[52] Badyal, Y., M.-L. Saboungi, D. Price, S. Shastri, D. Haeffner und A. Soper. Electron distribution in water. *J. Chem. Phys.* **112**, 9206–9208 (2000).

[53] Sprik, M. Hydrogen bonding and the static dielectric constant in liquid water. *J. Chem. Phys.* **95**, 6762–6769 (1991).

[54] Coutinho, K., R. Guedes, B. C. Cabral und S. Canuto. Electronic polarization of liquid water: converged Monte Carlo-quantum mechanics results for the multipole moments. *Chem. Phys. Lett.* **369**, 345–353 (2003).

[55] Tröster, P., K. Lorenzen, M. Schwörer und P. Tavan. Polarizable Water Models from Mixed Computational and Empirical Optimization. *J. Phys. Chem. B* **117**, 9486–9500 (2013).

[56] Tan, M.-L., J. R. Cendagorta und T. Ichiye. The molecular charge distribution, the hydration shell, and the unique properties of liquid water. *J. Chem. Phys.* **141**, 244504 (2014).

[57] Creighton, T. E. *Proteins*. W. H. Freeman and Company, San Francisco, 2. Auflage (1993).

[58] Schwörer, M., C. Wichmann und P. Tavan. A Polarizable QM/MM Approach to the Molecular Dynamics of Amide Groups Solvated in Water. *J. Chem. Phys.* **144**, 114504 (2016).

[59] Gaigeot, M. P., R. Vuilleumier, M. Sprik und D. Borgis. Infrared spectroscopy of N-methylacetamide revisited by ab initio molecular dynamics simulations. *J. Chem. Theory Comput.* **1**, 772–789 (2005).

[60] Ingrosso, F., G. Monard, M. Hamdi Farag, A. Bastida und M. F. Ruiz-López. Importance of Polarization and Charge Transfer Effects to Model the Infrared Spectra of Peptides in Solution. *J. Chem. Theory Comput.* **7**, 1840–1849 (2011).

[61] Wilson, E. B., J. C. Decius und P. C. Cross. *Molecular vibrations: the theory of infrared and Raman vibrational spectra*. Dover Publications, Inc. (1980).

[62] Moreau, C. und G. Douhéret. Thermodynamic and physical behaviour of water+ acetonitrile mixtures. Dielectric properties. *J. Chem. Thermodynamics* **8**, 403–410 (1976).

[63] Schultheis, V., R. Reichold, B. Schropp und P. Tavan. A Polarizable Force Field for Computing the Infrared Spectra of the Polypeptide Backbone. *J. Phys. Chem. B* **112**, 12217–12230 (2008).

[64] Pauling, L., R. B. Corey und H. R. Branson. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **37**, 205–211 (1951).

[65] Pauling, L. und R. B. Corey. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. USA* **37**, 251–256 (1951).

[66] Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey und M. L. Klein. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926–935 (1983).

[67] MacKerell, A., D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguven, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin und M. Karplus. All-Atom Empircal Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem.* **102**, 3586–3626 (1998).

[68] MacKerell, A. D., M. Feig und C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).

[69] Ramachandran, G. N., C. Ramakrishnan und V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).

[70] Feig, M. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *J. Chem. Theory Comput.* **4**, 1555–1564 (2008).

[71] Brooks, C. und D. A. Case. Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.* **93**, 2487–2502 (1993).

[72] Krimm, S. und J. Bandekar. Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem.* **38**, 181–364 (1986).

[73] Pelton, J. T. und L. R. McLean. Spectroscopic methods for analysis of protein secondary structure. *Anal. Biochem.* **277**, 167–176 (2000).

[74] Haris, P. I. und D. Chapman. The conformational analysis of peptides using Fourier transform IR spectroscopy. *Biopolymers* **37**, 251–263 (1995).

[75] Jackson, M., P. I. Haris und D. Chapman. Conformational transitions in poly (L-lysine): studies using Fourier transform infrared spectroscopy. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **998**, 75–79 (1989).

[76] Eaton, W. A., V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry und J. Hofrichter. Fast kinetics and mechanisms in protein folding 1. *Ann. Rev. Biophys. Biomol. Struct.* **29**, 327–359 (2000).

[77] Marx, D. und J. Hutter. *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press (2009).

[78] Born, M. und R. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys.* **389**, 457–484 (1927).

[79] Goldstein, H., C. P. Poole und J. L. Safko. *Klassische Mechanik*. Wiley-VCH, Berlin, 3. Auflage (2006).

[80] Verlet, L. Computer Experiments on classical fluids. (I) Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **165**, 201–214 (1967).

[81] Swope, W. C. und H. C. Anderson. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **76**, 637–649 (1982).

[82] Fließbach, T. *Quantenmechanik: Lehrbuch zur Theoretischen Physik III*. Springer, Berlin, 5. Auflage (2008).

[83] Ponder, J. W., D. A. Case *et al*. Force fields for protein simulations. *Advances in protein chemistry* **66**, 27–86 (2003).

[84] Rapaport, D. C. *The Art of Molecular Dynamics Simulation*. Cambridge university press, Cambridge, UK (2004).

[85] Guillot, B. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **101**, 219–260 (2002).

[86] Vega, C. und J. L. Abascal. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.* **13**, 19663–19688 (2011).

[87] Jones, J. E. On the determination of molecular fields. II. From the equation of state of a gas. In *Proc. R. Soc. London. Ser. A.*, Band 106, S. 463–477 (1924).

[88] Buckingham, R. A. und J. Corner. Tables of Second Virial and Low-Pressure Joule-Thomson Coefficients for Intermolecular Potentials with Exponential Repulsion. *Proc. R. Soc. London, Ser. A* **189**, 118–129 (1947).

[89] Jackson, J. D. *Klassische Elektrodynamik*. Walter de Gruyter, Berlin, New York, 4. Auflage (2006).

[90] Paricaud, P., M. Předota, A. Chialvo und P. Cummings. From Dimer to Condensed Phases at Extreme Conditions: Accurate Predictions of the Properties of Water by a Gaussian Charge Polarizable Model. *J. Chem. Phys.* **122**, 244511 (2005).

[91] Kiss, P. T. und A. Baranyai. A Systematic Development of a Polarizable Potential of Water. *J. Chem. Phys.* **138**, 204507 (2013).

[92] Bauer, S., G. Mathias und P. Tavan. Electrostatics of proteins in dielectric solvent continua. I. An accurate and efficient reaction field description. *J. Chem. Phys.* **140**, 104102 (2014).

[93] Yu, H. und W. F. Van Gunsteren. Accounting for Polarization in Molecular Simulation. *Comput. Phys. Commun.* **172**, 69–85 (2005).

[94] Schropp, B., C. Wichmann und P. Tavan. Spectroscopic Polarizable Force Field for Amide Groups in Polypeptides. *J. Phys. Chem. B* **114**, 6740–6750 (2010).

[95] Ahlström, P., A. Wallqvist, S. Engström und B. Jonsson. A Molecular-Dynamics Study of Polarizable Water. *Mol. Phys.* **68**, 563–581 (1989).

[96] Lamoureux, G., E. Harder, I. V. Vorobyov, B. Roux und A. D. MacKerell. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **418**, 245–249 (2006).

[97] Lifson, S. und A. Warshel. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules. *J. Chem. Phys.* **49**, 5116–5129 (1968).

[98] Lifson, S. und P. S. Stern. Born–Oppenheimer energy surfaces of similar molecules: Interrelations between bond lengths, bond angles, and frequencies of normal vibrations in alkanes. *J. Chem. Phys.* **77**, 4542–4550 (1982).

[99] Palmö, K., L.-O. Pietilä und S. Krimm. Construction of molecular mechanics energy functions by mathematical transformation of ab initio force fields and structures. *J. Comput. Chem.* **12**, 385–390 (1991).

[100] Mannfors, B., K. Palmo und S. Krimm. Spectroscopically determined force field for water dimer: physically enhanced treatment of hydrogen bonding in molecular mechanics energy functions. *J. Phys. Chem. A* **112**, 12667–12678 (2008).

[101] Warshel, A., M. Kato und A. Pisliakov. Polarizable force fields: history, test cases, and prospects. *J. Chem. Theory Comput.* **3**, 2034–2045 (2007).

[102] Lopes, P. E. M., B. Roux und A. D. Mackerell Jr. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability. Theory and applications. *Theoret. Chem. Acc.* **124**, 11–28 (2009).

[103] Cieplak, P., F.-Y. Dupradeau, Y. Duan und J. Wang. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **21**, 333102 (2009).

[104] Rick, S. W., S. J. Stuart und B. J. Berne. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **101**, 6141–6156 (1994).

[105] Anisimov, V. M., G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux und A. D. MacKerell. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.* **1**, 153–168 (2005).

[106] Applequist, J., J. R. Carl und K.-K. Fung. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **94**, 2952–2960 (1972).

[107] Russell, S. und A. Warshel. Calculations of electrostatic energies in proteins: The energetics of ionized groups in bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **185**, 389 – 404 (1985).

[108] Belle, D. V., I. Couplet, M. Prevost und S. J. Wodak. Calculations of electrostatic properties in proteins: Analysis of contributions from induced protein dipoles. *J. Mol. Biol.* **198**, 721 – 735 (1987).

[109] Lorenzen, K., C. Wichmann und P. Tavan. Including the Dispersion Attraction into Structure-Adapted Fast Multipole Expansions for MD Simulations. *J. Chem. Theory Comput.* **10**, 3244–3259 (2014).

[110] Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **59**, 341–350 (1981).

[111] Elking, D., T. Darden und R. J. Woods. Gaussian induced dipole polarization model. *J. Comput. Chem.* **28**, 1261–1274 (2007).

[112] Schwörer, M., B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan und G. Mathias. Coupling Density Functional Theory to Polarizable Force Fields for Efficient and Accurate Hamiltonian Molecular Dynamics Simulations. *J. Chem. Phys.* **138**, 244103 (2013).

[113] Niu, S., M.-L. Tan und T. Ichiye. The large quadrupole of water molecules. *J. Chem. Phys.* **134**, 134501 (2011).

[114] Tröster, P. und P. Tavan. The Microscopic Physical Cause for the Density Maximum of Liquid Water. *J. Phys. Chem. Lett.* **5**, 138–142 (2014).

[115] Andrés Cisneros, G., M. Karttunen, P. Ren und C. Sagui. Classical Electrostatics for Biomolecular Simulations. *Chem. Rev.* **114**, 779–814 (2013).

[116] Mathias, G., B. Egwolf, M. Nonella und P. Tavan. A Fast Multipole Method Combined with a Reaction Field for Long-range Electrostatics in Molecular Dynamics Simulations: The Effects of Truncation on the Properties of Water. *J. Chem. Phys.* **118**, 10847–10860 (2003).

[117] Darden, T. A., D. York und L. Pedersen. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

[118] Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee und L. G. Pedersen. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

[119] Luty, B. A., I. G. Tironi und W. F. van Gunsteren. Lattice-Sum Methods for Calculating Electrostatic Interactions in Molecular Simulations. *J. Chem. Phys.* **103**, 3014–3021 (1995).

[120] Appel, A. A. An Efficient Program for Many-body Simulation. *SIAM J. Sci. Stat. Comput.* **6**, 85–103 (1985).

[121] Barnes, J. und P. Hut. A Hierarchical $O(N \log N)$ Force-calculation Algorithm. *Nature* **324**, 446–449 (1986).

[122] Greengard, L. und V. Rokhlin. A Fast Algorithm for Particle Simulations. *J. Chem. Phys.* **73**, 325 – 348 (1987).

[123] Niedermeier, C. und P. Tavan. A Structure Adapted Multipole Method for Electrostatic Interactions in Protein Dynamics. *J. Chem. Phys.* **101**, 734–748 (1994).

[124] Niedermeier, C. und P. Tavan. Fast Version of the Structure Adapted Multipole Method — Efficient Calculation of Electrostatic Forces in Protein Dynamics. *Mol. Simul.* **17**, 57–66 (1996).

[125] Eichinger, M., H. Grubmüller, H. Heller und P. Tavan. FAMUSAMM: An Algorithm for Rapid Evaluation of Electrostatic Interactions in Molecular Dynamics Simulations. *J. Comput. Chem.* **18**, 1729–1749 (1997).

[126] Lorenzen, K., M. Schwörer, P. Tröster, S. Mates und P. Tavan. Optimizing the Accuracy and Efficiency of Fast Hierarchical Multipole Expansions for MD Simulations. *J. Chem. Theory Comput.* **8**, 3628–3636 (2012).

[127] Lorenzen, K., G. Mathias und P. Tavan. Linearly Scaling and Almost Hamiltonian Dielectric Continuum Molecular Dynamics Simulations through Fast Multipole Expansions. *J. Chem. Phys.* **143**, 184114 (2015).

[128] Martinetz, T. M., S. G. Berkovich und K. J. Schulten. 'Neural-gas' network for vector quantization and its application to time-series prediction. *IEEE Transactions on Neural Networks* **4**, 558–569 (1993).

[129] Tröster, P., K. Lorenzen und P. Tavan. Polarizable Six-Point Water Models from Computational and Empirical Optimization. *J. Phys. Chem. B* **118**, 1589 (2014).

[130] Mathias, G. und P. Tavan. Angular resolution and range of dipole–dipole correlations in water. *J. Chem. Phys.* **120**, 4393–4403 (2004).

[131] Bauer, S., P. Tavan und G. Mathias. Electrostatics of Proteins in Dielectric Solvent Continua. II. Hamiltonian Reaction Field Dynamics. *J. Chem. Phys.* **140**, 104103 (2014).

[132] Bauer, S., P. Tavan und G. Mathias. Exploring Hamiltonian dielectric solvent molecular dynamics. *Chem. Phys. Lett.* **612**, 20–24 (2014).

[133] Mennucci, B. und J. M. Martínez. How to model solvation of peptides? Insights from a quantum-mechanical and molecular dynamics study of N-methylacetamide. 1. Geometries, infrared, and ultraviolet spectra in water. *J. Phys. Chem. B* **109**, 9818–9829 (2005).

[134] Puchner, E. M., A. Alexandrovich, A. L. Kho, U. Hensen, L. V. Schäfer, B. Brandmeier, F. Gräter, H. Grubmüller, H. E. Gaub und M. Gautel. Mechanoenzymatics of titin kinase. *Proc. Natl. Acad. Sci. USA* **105**, 13385–13390 (2008).

[135] Freddolino, P. L., A. S. Arkhipov, S. B. Larson, A. McPherson und K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **14**, 437–449 (2006).

[136] Gunsteren, W. F., D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt und H. B. Yu. Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chem. Int. Ed.* **45**, 4064–4092 (2006).

[137] Babitzki, G., R. Denschlag und P. Tavan. Polarization Effects Stabilize Bacteriorhodopsin's Chromophore Binding Pocket: A Molecular Dynamics Study. *J. Phys. Chem. B* **113**, 10483–10495 (2009).

[138] Demerdash, O., E.-H. Yap und T. Head-Gordon. Advanced potential energy surfaces for condensed phase simulation. *Ann. Rev Phys. Chem.* **65**, 149–174 (2014).

[139] Parr, R. G. und W. Yang. *Density-functional theory of atoms and molecules*. Oxford University Press, New York (1989).

[140] Szabo, A. und N. S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation (1989).

[141] Hohenberg, P. und W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864–B870 (1964).

[142] Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **140**, 18A301 (2014).

[143] Kohn, W. und L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, A1133–A1138 (1965).

[144] Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).

[145] Becke, A. D. Density-functional Exchange-energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).

[146] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).

[147] Stephens, P., F. Devlin, C. Chabalowski und M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry* **98**, 11623–11627 (1994).

[148] Perdew, J. und W. Yue. Accurate and Simple Density Functional for the Electronic Exchange Energy: Generalized Gradient Approximation. *Phys. Rev. B* **33**, 8800–8802 (1986).

[149] Lee, C., W. Yang und R. G. Parr. Development of the Colle-Salvetti Correlation-Energy Formula Into a Functional of the Electron Density. *Phys. Rev. B* **37**, 785–789 (1988).

[150] Hehre, W. J., R. F. Stewart und J. A. Pople. Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals. *J. Chem. Phys.* **51**, 2657–2664 (1969).

[151] Dunning Jr, T. H. und P. J. Hay. Gaussian basis sets for molecular calculations. In *Methods of electronic structure theory*, S. 1–27. Springer (1977).

[152] Hutter, J., A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman und M. Parrinello. CPMD: Car–Parinello Molecular Dynamics, version 3.17.1. © IBM Corp 1990–2008 and MPI für Festkörperforschung Stuttgart 1997–2001.

[153] Lippert, G., J. Hutter und M. Parrinello. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **92**, 477–487 (1997).

[154] VandeVondele, J., M. Krack, F. Mohamed, M. Parrinello, T. Chassaing und J. Hutter. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **167**, 103–128 (2005).

[155] Hutter, J., M. Iannuzzi, F. Schiffmann und J. VandeVondele. CP2K: atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci.* **4**, 15–25 (2014).

[156] Troullier, N. und J. L. Martins. Efficient Pseudopotentials for Plane-wave Calculations. *Phys. Rev. B* **43**, 1993–2005 (1991).

[157] Car, R. und M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).

[158] Nobelprize.org. Press Release: The 1998 Nobel Prize in Chemistry. `http://nobelprize.org/nobel_prizes/chemistry/laureates/1998/press.html`, besucht am 9. Februar 2016.

[159] Nonella, M. und P. Tavan. An unscaled quantum mechanical harmonic force field for p-benzoquinone. *Chem. Phys.* **199**, 19–32 (1995).

[160] Bytheway, I. und M. W. Wong. The prediction of vibrational frequencies of inorganic molecules using density functional theory. *Chem. Phys. Lett.* **282**, 219–226 (1998).

[161] Neugebauer, J. und B. A. Hess. Fundamental Vibrational Frequencies of Small Polyatomic Molecules from Density-Functional Calculations and Vibrational Perturbation Theory. *J. Chem. Phys.* **118**, 7215–7225 (2003).

[162] Mantz, Y. A., H. Gerard, R. Iftimie und G. J. Martyna. Ab initio and empirical model MD simulation studies of solvent effects on the properties of N-methylacetamide along a cis-trans isomerization pathway. *J. Phys. Chem. B* **110**, 13523–13538 (2006).

[163] Gaigeot, M.-P. Unravelling the conformational dynamics of the aqueous alanine dipeptide with first-principle molecular dynamics. *J. Phys. Chem. B* **113**, 10059–62 (2009).

[164] VandeVondele, J., P. Tröster, P. Tavan und G. Mathias. Vibrational spectra of phosphate ions in aqueous solution probed by first-principles molecular dynamics. *J. Phys. Chem. A* **116**, 2466–74 (2012).

[165] Laasonen, K., M. Sprik, M. Parrinello und R. Car. "Ab initio" liquid water. *J. Chem. Phys.* **99**, 9080–9089 (1993).

[166] Sprik, M., J. Hutter und M. Parrinello. Ab initio molecular dynamics simulation of liquid water: Comparison of three gradient-corrected density functionals. *J. Chem. Phys.* **105**, 1142–1152 (1996).

[167] Silvestrelli, P. L., M. Bernasconi und M. Parrinello. Ab initio infrared spectrum of liquid water. *Chem. Phys. Lett.* **277**, 478–482 (1997).

[168] VandeVondele, J., F. Mohamed, M. Krack, J. Hutter, M. Sprik und M. Parrinello. The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *J. Chem. Phys.* **122**, 014515 (2005).

[169] Todorova, T., A. P. Seitsonen, J. Hutter, I.-F. W. Kuo und C. J. Mundy. Molecular dynamics simulation of liquid water: hybrid density functionals. *J. Phys. Chem. B* **110**, 3685–3691 (2006).

[170] Gao, J. und D. G. Truhlar. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **53**, 467–505 (2002).

[171] Lin, H. und D. G. Truhlar. QM/MM: what have we learned, where are we, and where do we go from here? *Theoret. Chem. Acc.* **117**, 185–199 (2006).

[172] Senn, H. M. und W. Thiel. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **48**, 1198–1229 (2009).

[173] Mennucci, B. Modeling environment effects on spectroscopies through QM/classical models. *Phys. Chem. Chem. Phys.* **15**, 6583–94 (2013).

[174] Schwörer, M. und G. Mathias. Modeling Biomolecular Solvation Effects by Hybrid QM/MM Methods. In G. Sutmann, J. Grotendorst, G. Gompper und D. Marx (Herausgeber), *Computational Trends in Solvation and Transport in Liquids*, S. 315–332. Schriften des Forschungszentrums Jülich, IAS Series 28 (2015).

[175] Eichinger, M., H. Grubmüller und H. Heller. *User Manual for EGO_VIII, Release 2.0*. Theoretische Biophysik, Institut für Medizinische Optik, Ludwig-Maximilians-Universität, Oettingenstr. 67, D-80585 München (1995).

[176] Eichinger, M. *Berechnung molekularer Eigenschaften in komplexer Lösungsmittelumgebung: Dichtefunktionaltheorie kombiniert mit einem Molekularmechanik-Kraftfeld*. Dissertation, Ludwig-Maximilians Universität München, Lehrstuhl für BioMolekulare Optik, AG Theoretische Biophysik (1999).

[177] Nonella, M., G. Mathias und P. Tavan. Infrared Spectrum of *p*-Benzoquinone in Water Obtained from a QM/MM Hybrid Molecular Dynamics Simulation. *J. Phys. Chem. A* **107**, 8638–8647 (2003).

[178] Klähn, M., J. Schlitter und K. Gerwert. Theoretical IR spectroscopy based on QM/MM calculations provides changes in charge distribution, bond lengths, and bond angles of the GTP ligand induced by the Ras-protein. *Biophys. J.* **88**, 3829–3844 (2005).

[179] te Heesen, H., K. Gerwert und J. Schlitter. Role of the arginine finger in Ras·RasGAP revealed by QM/MM calculations. *FEBS lett.* **581**, 5677–5684 (2007).

[180] Schmitz, M. und P. Tavan. Vibrational Spectra from Atomic Fluctuations in Dynamics Simulations. I. Theory, Limitations, and a Sample Application. *J. Chem. Phys.* **121**, 12233–12246 (2004).

[181] Schmitz, M. und P. Tavan. Vibrational Spectra From Atomic Fluctuations in Dynamics Simulations. II. Solvent-induced Frequency Fluctuations at Femtosecond Time Resolution. *J. Chem. Phys.* **121**, 12247–12258 (2004).

[182] Rieff, B., G. Mathias, S. Bauer und P. Tavan. Density functional theory combined with molecular mechanics: the infrared spectra of flavin in solution. *Photochem. Photobiol.* **87**, 511–523 (2011).

[183] Schropp, B. und P. Tavan. The Polarizability of point-polarizable Water Models: Density Funtional Theory / Molecular Mechanics Results. *J. Phys. Chem. B* **112**, 6233–6240 (2008).

[184] Borysow, J., M. Moraldi und L. Frommhold. The collision induced spectroscopies: Concerning the desymmetrization of classical line shape. *Mol. Phys.* **56**, 913–922 (1985).

[185] Mathias, G. und M. Baer. Generalized normal coordinates for the vibrational analysis of molecular dynamics simulations. *J. Chem. Theory Comput.* **7**, 2028–2039 (2011).

[186] Mathias, G., S. D. Ivanov, A. Witt, M. D. Baer und D. Marx. Infrared Spectroscopy of Fluxional Molecules from (ab Initio) Molecular Dynamics: Resolving Large-Amplitude Motion, Multiple Conformations, and Permutational Symmetries. *J. Chem. Theory Comput.* **8**, 224–234 (2011).

[187] Breitenfeld, B. *Kopplung des Dichtefunktionaltheorie-Programms CPMD an ein polarisierbares Kraftfeld: Implementierung und Tests*. Diplomarbeit, Ludwig-Maximilians Universität München, Lehrstuhl für BioMolekulare Optik, AG Theoretische Biophysik (2010).

[188] Laio, A., J. VandeVondele und U. Rothlisberger. A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car-Parrinello Molecular Dynamics Simulations. *J. Chem. Phys.* **116**, 6941–6947 (2002).

[189] Laino, T., F. Mohamed, A. Laio und M. Parrinello. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J. Chem. Theory Comput.* **1**, 1176–1184 (2005).

[190] Carfi, D. The pointwise Hellmann-Feynman theorem. *AAPP Physical, Mathematical, and Natural Sciences* **88** (2010).

[191] Schwörer, M. *Quantum-classical coupling of polarizable molecular mechanics force fields*. Masterarbeit, Ludwig-Maximilians Universität München, Lehrstuhl für BioMolekulare Optik, AG Theoretische Biophysik (2012).

[192] IPHIGENIE ist unter der GPL Lizenz frei downloadbar unter `http://sourceforge.net/projects/iphigenie`.

[193] Schwörer, M., K. Lorenzen, G. Mathias und P. Tavan. Utilizing Fast Multipole Expansions for Efficient and Accurate Quantum-Classical Molecular Dynamics Simulations. *J. Chem. Phys.* **142**, 104108 (2015).

[194] Schwörer, M., C. Wichmann, E. Gawehn und G. Mathias. Simulated Solute Tempering in Fully Polarizable Hybrid QM/MM Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **12**, 992–999 (2016).

[195] Denschlag, R., M. Lingenheil, P. Tavan und G. Mathias. Simulated solute tempering. *J. Chem. Theory Comput.* **5**, 2847–2857 (2009).

[196] Hansen, H. S. und P. H. Hünenberger. Using the local elevation method to construct optimized umbrella sampling potentials: calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. *J. Comput. Chem.* **31**, 1–23 (2010).

[197] Mitsutake, A., Y. Sugita und Y. Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **60**, 96–123 (2001).

[198] Lyubartsev, A., A. Martsinovski, S. Shevkunov und P. Vorontsov-Velyaminov. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.* **96**, 1776–1783 (1992).

[199] Marinari, E. und G. Parisi. Simulated tempering: a new Monte Carlo scheme. *Europhys. Lett.* **19**, 451 (1992).

[200] Swendsen, R. H. und J.-S. Wang. Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* **57**, 2607–2609 (1986).

[201] Sugita, Y. und Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).

[202] Liu, P., B. Kim, R. A. Friesner und B. J. Berne. Replica Exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA* **102**, 13749–13754 (2005).

[203] Wang, L., R. A. Friesner und B. Berne. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. A* **115**, 9431–9438 (2011).

[204] Laio, A., J. VandeVondele und U. Rothlisberger. D-RESP: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. *J. Phys. Chem. B* **106**, 7300–7307 (2002).

[205] Klimeš, J. und A. Michaelides. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **137**, 120901 (2012).

[206] Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).

[207] Cochran, A. G., N. J. Skelton und M. A. Starovasnik. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. USA* **98**, 5578–5583 (2001).

Erklärung

Ich versichere hiermit, dass die Dissertation von mir selbständig, ohne unerlaubte Hilfe angefertigt worden ist.

Weiterhin erkläre ich hiermit, dass ich mich anderweitig keiner Doktorprüfung ohne Erfolg unterzogen habe.

Ferner erkläre ich hiermit, dass ich diese Dissertation keiner anderen Prüfungskommission vorgelegt habe.

München, den 23.03.2016

Magnus Schwörer

Danksagung

Zuguterletzt möchte ich allen, die zum Erfolg dieser Arbeit beigetragen haben, an dieser Stelle ganz herzlich danken.

Prof. Dr. Paul Tavan danke ich dafür, dass er mich — ungeplanterweise — als Doktorand aufgenommen hat und meine Arbeit bis weit in seinen Ruhestand hinein betreut hat. Seine Unterstützung trug wesentlich zur erfolgreichen Bearbeitung dieses Themas bei, auch in wissenschaftlicher Hinsicht hab ich viel von ihm gelernt.

Auch Gerald Mathias sei für sein immer offenes Ohr und seine vielen guten Ideen gedankt. Die trotz aller Effizienzsteigerungen doch immer noch recht aufwändigen Test- und Produktivrechnungen wären ohne die von ihm organisierten Rechenzeitressourcen sehr viel schwieriger zu realisieren gewesen.

In diesem Zusammenhang danke ich dem LRZ in Person von Helmut Satzger, der mir ein wertvoller Ansprechpartner war, sowie Nicolay Hammer, mit dessen Hilfe ich erfolgreich ausgedehnte (Test-)Rechnungen auf SuperMUC Phase 2 durchführen konnte. Danke auch an Michael Stübiger vom Institut für Informatik der LMU für den Zugang zum CIP-Cluster und den umfassenden Support.

Meinen Arbeitsgruppenkollegen Christoph, Erik, Konstantin, Maxim, Nico und Sebastian danke ich für die angenehme Zusammenarbeit, die sich durch eine gesunde Mischung aus wissenschaftlicher Diskussion, konkreter Hilfestellung und Ablenkung auszeichnete. Besonders Christoph war in schwerer See ein zuverlässiger Bürokollege. Auch allen anderen Mitarbeitern am BMO, die zur angenehmen Arbeitsatmosphäre beigetragen haben, sei gedankt.

Bei meinem Freundeskreis möchte ich mich für die zahlreichen arbeitsbedingten Absenzen entschuldigen. Meinem früheren Kommilitonen Sebastian danke ich für eine wichtige Diskussion.

Meinen Eltern gebührt Dank für ihre Unterstützung, durch die ich mich auch gegen Ende der Promotion voll auf die Arbeit konzentrieren konnte.

Liebe Eva, danke dass du meine schwankenden Launen ertragen hast, und mir auch in schweren Zeiten immer ein großer Rückhalt warst!